The international journal of science / 6 April 2023

nature



TUMOUR PROMOTION

Air pollution drives lung cancer in non-smokers

Treatment boost
The push to create a new generation of Alzheimer's therapies

Warming waters
Define marine heatwaves to help coastal communities

Dark of the Moon
Lunar eclipse records shed light on medieval volcanic activity

VISIT…

LANZAROTE Caliente.COM

Nature.2023.04.08

[Sat, 08 Apr 2023]

- **This Week**
- **News in Focus**
- **Opinion**
- **Work**
- **Research**

This Week

- **European backsliding on electric vehicles is bad news for the climate** [04 Apr 2023]
  • Attempts to put a brake on the transition to electrification and allow 'climate neutral' fuels after 2035 ignore the science — what's needed is policy clarity now.
- **The WHO at 75: what doesn't kill you makes you stronger** [05 Apr 2023]
  • The World Health Organization is emerging from the peak of the pandemic bruised. Its member states must get back to prioritizing universal health care.
- **Carbon dioxide removal is not a current climate solution — we need to change the narrative** [04 Apr 2023]
  • Drastically reduce emissions first, or carbon dioxide removal will be next to useless.

- EDITORIAL
- 04 April 2023

European backsliding on electric vehicles is bad news for the climate

Attempts to put a brake on the transition to electrification and allow ‘climate neutral’ fuels after 2035 ignore the science — what’s needed is policy clarity now.



The future of mobility needs to be electric.Credit: Shutterstock

Worldwide, the planes, trains and automobiles we use to get around pumped around 7.7 gigatonnes of carbon dioxide into the atmosphere in 2021, one-fifth of all anthropogenic emissions. Some three-quarters of transport emissions came from just one source — the exhausts of road vehicles.

Converting road transport to run on green energy would be a huge step towards achieving net zero emissions by mid-century, a change needed if we are to limit global warming to 'safe' levels. This is why policymakers have been nudging car makers to accelerate efforts to bring an end to the manufacture of vehicles fitted with an internal combustion engine. It's a no-brainer. In the European Union, at least, it seemed that the two sides were strapped in, ready to reach that destination by 2035.

However, the past few weeks have seen the European Commission embroiled in a row with Germany, Italy and some other EU members over implementation of the 2035 deadline. This has been resolved, but only through a concession to Germany's powerful automotive industry. New cars with internal combustion engines can continue to be sold after 2035, provided the engines use carbon-neutral fuels instead of diesel, petrol or compressed and liquefied gases. These are climate-damaging moves from a region that has so far led the world in policies for decarbonizing transport.



How the hydrogen revolution can help save the planet — and how it can't

The problem lies in the phrase 'carbon-neutral fuels'. These fuels rely either on inputs such as 'green' hydrogen, which is made by splitting water using renewable electricity, or on feedstocks such as biomass. The technologies used to make these fuels are inefficient, expensive and untested at scale. Moreover, claims of climate neutrality — based on the idea that the CO_2

emitted by their combustion was absorbed relatively recently from the biosphere, or that CO_2 produced during their manufacture was prevented from entering the atmosphere — are questionable.

The capacity to make green hydrogen is severely limited, and any expansion should be used to power sectors such as heavy industry, for which viable decarbonization alternatives are not yet available. Meanwhile, the use of biomass creates incentives to harvest wood and divert agricultural land to grow energy crops, regardless of the consequences for land as a carbon sink or for biodiversity.

It’s clear why some in the automotive industry want to keep the internal combustion engine alive. The idea is attractive to short-sighted policymakers, too, because it reduces the need to plan the roll out of charging infrastructure, to worry about grid capacity, and to teach people the skills to build and maintain different technologies. The research community must be equally clear in underlining why this is a false economy. There is only one proven viable, scalable and technologically ripe scheme for decarbonizing personal road transport. That is electrification.

Not all car makers want to delay. Many understand that the transition to electric vehicles will take time, and want to get on with transforming their businesses. They want policy certainty and continuity from governments to allow them to get down to business. Last year’s COP27 climate conference in Sharm El-Sheikh, Egypt, saw the launch of the Accelerating to Zero Coalition to drive the global transition to new electric cars and vans by 2035 in “leading markets”, meaning high-income countries, and globally by 2040. Its more than 200 signatories include 14 car manufacturers, among them household names such as Ford, General Motors, Mercedes-Benz and Volvo Cars, and the governments of more than 40 countries.



Overhyping hydrogen as a fuel risks endangering net-zero goals

But the absentees are also notable. They include some of the world's most prominent motor manufacturers — Toyota, Volkswagen, Honda, Hyundai and Kia. Also absent are the governments of some of the biggest car-producing countries — China, Japan, South Korea and Germany.

If the electric-vehicle transition is further delayed, there are likely to be cascading effects elsewhere that will ultimately put a brake on global decarbonization. The demand for personal powered mobility is increasing in low- and middle-income countries. In Asia alone, cars are projected to account for more than 40% of trips taken in 2050, up from 28% in 2015. On the basis of current trends, there will be three billion cars and vans on the road globally in 2050, up from one billion now — another reason to accelerate the transition to electric vehicles worldwide.

For the decarbonization of road transport to occur, the world will need what the Global Fuel Economy Initiative, a partnership on fuel economy and efficiency, called a "radical policy framework" (see go.nature.com/4381wvk). That means the removal of fossil-fuel subsidies and the mobilization of both public and private investment for the development of electric vehicles and their attendant charging infrastructure. It means tying the development of that infrastructure to renewable-energy-generation systems, while ensuring that supply chains are sustainable and

providing recycling facilities for battery materials. And it means an international agreement must be reached on standards, so that the introduction of cleaner vehicles in one part of the world doesn't mean old bangers being shipped off to pollute the environment elsewhere.

All of this is doable. But the growing global demand for personal mobility means a truly green transport transition will happen only by addressing another factor. Technological innovation will take us only so far: behavioural change is also needed. Alongside a cogent, evidence-based strategy to develop electric vehicles and displace fossil fuels, we must plan and redesign urban environments around the world to encourage active transport — walking and cycling — rather than driving. That surely is the best route to a cleaner, healthier world.

Nature **616**, 7 (2023)

doi: https://doi.org/10.1038/d41586-023-00951-z

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00951-z

- EDITORIAL
- 05 April 2023

The WHO at 75: what doesn't kill you makes you stronger

The World Health Organization is emerging from the peak of the pandemic bruised. Its member states must get back to prioritizing universal health care.



The World Health Organization's smallpox eradication campaign (seen here in India) has been one of its biggest successes.Credit: Tambarahalli S. Satyan/WHO

There were times during the first year of the COVID-19 pandemic when the World Health Organization (WHO) must have felt as if it had been plunged into the middle of a heavyweight boxing match between China and the United States. China stopped its virologists and epidemiologists

collaborating with their international colleagues just as researchers needed to share what they knew to understand the new virus. The United States, under then-president Donald Trump, blamed the WHO for being too close to China and attempted to punish the agency by announcing plans to withdraw funding (later reversed by the administration of current President Joe Biden).

But the WHO is no stranger to big power conflicts. Throughout its 75-year existence, it has been surrounded by international tensions, most notably the cold war between the United States and the Soviet Union (E. Fee *et al. Am. J. Public Health* **106**, 1912–1917; 2016). The US Congress approved US membership of the WHO on the condition that it could withdraw with a year's notice. The Soviet Union did pull away between 1949 and 1955.

When thinking about the WHO's 75 years, it's worth remembering the time and circumstances of its creation. In the aftermath of the Second World War, the newly established United Nations and its specialized agencies, including the WHO, were designed to future-proof the world from another global conflict. Around 80 million people died during the two world wars, many from famine or disease.



The WHO deserves more money for its core mission — and more respect

The WHO's founding constitution states unequivocally: "Health is a state of complete physical, mental and social well-being and not merely the absence

of disease or infirmity. The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition."

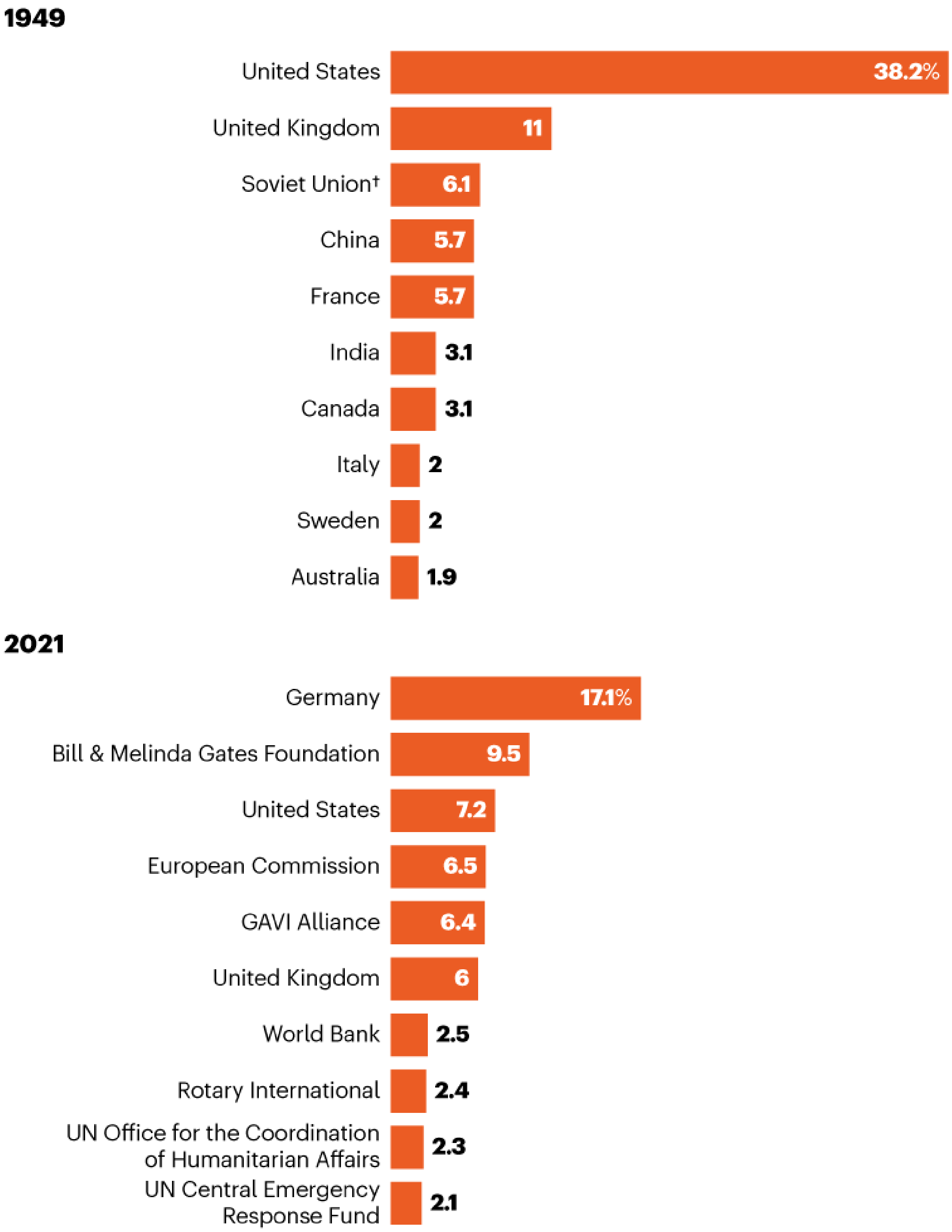
And yet, the agency's creators chose not to prioritize robust systems of universal health care that would meet these goals. This absent focus is one factor in why infectious diseases continue to impact populations in low- and middle-income countries. The eradication of smallpox in 1980 was a big win. But for other diseases, the agency and its donors have been unable to reach targets, including in the elimination of HIV and AIDS, malaria and tuberculosis.

The WHO does, however, have a consistent record for establishing itself as the go-to organization for setting global standards for the efficacy, safety and quality of vaccines and medicines. As we have seen during the pandemic, the agency is central to alerting the world to new infectious diseases, helped in no small measure by the revolution in biomedicine and health data, especially genomics.

How the WHO operates and sets its priorities is also influenced by the economic-policy priorities of its member states. The agency was created by a handful of powers, many of which are still significant donors (see 'Fallen ideals'). Contributions from the United States (including the Bill & Melinda Gates Foundation in Seattle, Washington) and European nations accounted for more than half of the agency's US$7.6-billion income in 2021.

FALLEN IDEALS

The US government contributed nearly 40% of the World Health Organization's US$5-million budget at the end of 1949. By contrast, the United States provided around 7% of the agency's annual 'voluntary' contributions* to the end of 2021.



*Voluntary funding from member states as well as from sources such as philanthropic foundations amounted to more than 80% of the WHO's $7.6-billion budget for that year. †The Soviet Union pledged around $300,000 to the WHO before it withdrew from the agency in 1949, rejoining in 1955.

©nature

Source: WHO

Boosting innovation and competitiveness has been a big priority, and this has come at the expense of knowledge sharing. In 1980, the United States enacted the Bayh–Dole Act allowing universities to commercialize inventions from federally funded research. This led to restrictions on what researchers could and couldn't share. In 1995, the World Trade Organization established rules on intellectual property that have also made it difficult to share the knowledge needed to develop COVID-19 vaccines and therapies quickly and at scale. India and South Africa mounted a brave campaign (*Nature* was among its supporters) to have COVID-19-related intellectual-property rights temporarily waived; but the odds were stacked against them.

The failure of COVAX, a scheme to provide vulnerable populations with COVID-19 vaccines, also prompted reflection on what is wrong with the architecture of global public health. COVAX struggled because governments in a small number of high-income countries over-ordered (and overpaid for) vaccines, leaving little for the rest of the world.



Global pandemic treaty: what we must learn from climate-change errors

But the tide could be turning. African countries are determined to establish vaccine manufacturing on the continent. Knowledge sharing and vaccine

equity are also at the heart of the draft text for a pandemic treaty that is currently being negotiated at the WHO.

In the face of scepticism from some of the WHO's biggest donors, in 2021, the organization created a facility that offers the technology for messenger RNA vaccines to low-income countries, through an initiative created by former chief scientist Soumya Swaminathan and backed by director-general Tedros Adhanom Ghebreyesus.

Many continue to argue that strengthening health systems, sharing knowledge, and preventing and controlling diseases is principally the job of governments. The corollary is that the WHO should take part, but not lead this work. But it shouldn't be a case of having to choose. In some of the poorest parts of the world, the WHO is a second health ministry, running (or funding) primary-care clinics, delivering vaccinations and collecting data. In conflict zones, it is one of the few trusted organizations that can do all of these, as the world saw recently in northwest Syria.

Today, the WHO faces enormous challenges. Health systems are reeling amid the pandemic. COVID-19 has not only killed nearly seven million people, but also set back global-health targets — from reducing global maternal mortality to achieving universal health coverage.

The WHO has come through the past few years bruised, but stronger and wiser. Member states must get back to prioritizing universal health and supporting nations less able to achieve this on their own. They must prioritize equity, access to knowledge and rediscovering their founding ideals, so the world is better prepared for the next emergency.

Nature **616**, 8 (2023)

doi: https://doi.org/10.1038/d41586-023-00952-y

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00952-y

- WORLD VIEW
- 04 April 2023

Carbon dioxide removal is not a current climate solution — we need to change the narrative



Drastically reduce emissions first, or carbon dioxide removal will be next to useless.

- David T. Ho

Carbon dioxide removal (CDR) is what puts the ‘net’ into ‘net zero emissions’. All pathways to limit global warming to 1.5–2 °C above pre-industrial levels that have been assessed by the Intergovernmental Panel on Climate Change require rapid decarbonization to start now. But they also require the removal of CO_2 from the atmosphere because we won’t be able to eliminate carbon emissions entirely on the required time scales. ‘Hard to abate’ sectors such as aviation and shipping will remain large sources of

greenhouse gases even in the most optimistic scenarios. Residual emissions will mean that we cannot achieve a zero-emissions goal, and we will need CDR to reach a net-zero target. Historically, this has meant planting or maintaining trees, but removing CO_2 from the atmosphere and storing it in the ground, the ocean or in products will be more durable.

However, businesses are springing up that promise various CDR techniques as climate solutions for today. Others are enthusiastically buying carbon credits — essentially, investments in planting trees, or other future CDR capacity — as part of their current decarbonization commitments.

I have spent my career studying the natural carbon cycle and, in recent years, developing methods for checking that CDR works. I have scrutinized dozens of proposals, and I was a reviewer for the US$100-million XPRIZE Carbon Removal competition funded by the Musk Foundation. I don't deny the need to develop CDR methods over the longer term. And I welcome governments committing much-needed resources to this area. After some small-scale demonstrations of 'direct air capture' (DAC) technology, which suck CO_2 out of the atmosphere by chemical means, the 2022 US Bipartisan Infrastructure Law has devoted $3.5 billion to developing four DAC hubs. But it's clear to me that deploying them to remove CO_2 from the atmosphere is pointless until society has almost completely eliminated its polluting activities.

Time travel

To understand why, think of CDR as a time machine. Take the proposed US DAC hubs, for example. Each facility is eventually expected to extract one million tonnes of CO_2 each year.



[Carbon capture nets 2 billion tonnes of CO2 each year — but it's not enough]

In 2022, the world emitted 40.5 billion tonnes of CO_2 (P. Friedlingstein *et al. Earth Syst. Sci. Data* **14**, 4811–4900; 2022). At that rate, for every year of operation at its full potential, each hub would take the atmosphere back in time by almost 13 minutes, but in the time it took to remove those 13 minutes of CO_2, the world would have spewed another full year of CO_2 into the atmosphere.

Meanwhile, if everyone on Earth planted a tree — 8 billion trees — it would take us back in time by about 43 hours every year, once the trees had matured.

The time-machine analogy reveals just how futile CDR currently is.

We have to shift the narrative as a matter of urgency. Money is going to flood into climate solutions over the next few years, and we need to direct it well. We must stop talking about deploying CDR as a solution today, when emissions remain high — as if it somehow replaces radical, immediate emission cuts.

Back to the future

It's a different story if we successfully decarbonize enough in the coming 20–30 years.

If we reduced emissions to around 10% of current levels — 4 billion tonnes of CO_2 a year — a DAC plant capable of removing one million tonnes would be a time machine taking us back just over 2 hours instead of 13 minutes. At that point, it would take 4,000 facilities to reach net zero in any given year, presuming they were fully powered by renewable energy.



Net-zero emissions targets are vague: three ways to fix

In reality, residual emissions will probably be 18% of our current total (H. J. Buck *et al. Nature Clim. Change* https://doi.org/j4jg; 2023), so we will have to scale up CDR substantially to reach net zero. Still, it might be feasible to build 7,290 DAC hubs or deploy other CDR technology.

In the meantime, research is needed to seek CDR methods that minimize land use and energy consumption, and can be scaled up rapidly and cheaply. Doing that now is essential, so that we have the technology available in the future, when it will be effective, and when it can be used to remove legacy emissions to address intergenerational justice.

Not all techniques that work in the laboratory will work in the real world. Some might come with heavy costs for biodiversity and the environment.

Developing methods to verify that CDR works is a major challenge. It will be many years before we have the science to tell us which methods work and whether they harm or benefit the environment.

Humanity has never removed an atmospheric pollutant at a global, continental or, even, regional scale — we have only ever shut down the source and let nature do the clearing up. This is the case for chlorofluorocarbons and stratospheric ozone destruction, for sulfur dioxide and acid rain, and for sulfur and nitrogen oxides and photochemical smog. We must be prepared for CDR to be a failure, leaving us to rely on the environment to stabilize atmospheric CO_2 over thousands of years. This is another argument for rapid decarbonization.

The scale of the challenge is immense. We must slow the carbon clock to a crawl before we can turn it back.

Nature **616**, 9 (2023)

doi: https://doi.org/10.1038/d41586-023-00953-x

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00953-x

News in Focus

- **COVID-origins report sparks debate over major genome hub GISAID** [29 Mar 2023]
  • GISAID revoked researchers' access following the report, sparking discussion about findings based on data found in online repositories.
- **Researchers in Germany protest proposed postdoc rule change** [29 Mar 2023]
  • A potential legal reform is reigniting debate about a time limit on fixed-term postdoc contracts.
- **How green is your research? These scientists are cutting their carbon footprints** [28 Mar 2023]
  • As the effects of climate change grow, scientists from fields spanning astronomy to biology are trying to decarbonize their research.
- **Antibody-patent row could have far-reaching impact on biotech** [28 Mar 2023]
  • The results of a US case could dictate how broad patents are, and have knock-on effects for those developing drugs.
- **JWST gets best view yet of planet in hotly pursued star system** [27 Mar 2023]
  • Telescope didn't spot an atmosphere on TRAPPIST-1b, but has kicked off a new era in understanding planetary evolution.
- **'Astonishing' molecular syringe ferries proteins into human cells** [29 Mar 2023]
  • Technique borrowed from nature, and honed using artificial intelligence, could spur the development of better drug-delivery systems.
- **Facelift for T. rex: analysis suggests teeth were covered by thin lips** [30 Mar 2023]
  • Crocodiles and Komodo dragons provide evidence to support the idea of a scaly cover over the teeth of dinosaur Tyrannosaurus rex.
- **How to convince a politician: a science adviser's lessons from the pandemic** [30 Mar 2023]

• Patrick Vallance, who advised the UK government during the height of COVID, says that scientific officials should know four key things when guiding policymakers.

- **Diversity in German science: researchers push for missing ethnicity data** [05 Apr 2023]
  • The European country is one of several reassessing its cultural unease with collecting information on scientists' race and ethnicity.
- **Conquering Alzheimer's: a look at the therapies of the future** [04 Apr 2023]
  • Researchers are looking to drug combinations, vaccines and gene therapy as they forge the next generation of treatments for the condition.

- NEWS
- 29 March 2023
- Update 31 March 2023
- Correction 31 March 2023

COVID-origins report sparks debate over major genome hub GISAID

GISAID revoked researchers’ access following the report, sparking discussion about findings based on data found in online repositories.

- Smriti Mallapaty



The use of data in public repositories is under debate after a temporary suspension of 19 authors.Jackie Niam/iStock/Getty

On 20 March, an international team of researchers posted a headline-grabbing online report[1] describing genomic data found in swabs taken from the market in Wuhan, China, that has been linked to the origins of the COVID-19 pandemic. The data were gathered by a separate team in China and hosted on a major online database, the Global Initiative on Sharing Avian Influenza Data (GISAID). Hours later, GISAID temporarily revoked the authors' access.

GISAID said that the authors of the report, posted to the online platform Zenodo, had not done enough to protect the interests of the team in China — the 'data generators' — and had therefore violated its rules.

The authors of the Zenodo report — an international team of 19 scientists, many of whom have published papers on the origins of SARS-CoV-2 — deny this, saying that they had attempted to collaborate with the team in China, and noting that their new report acknowledges that team.

"I feel that we're being punished while abiding by both the letter and the spirit of their rules," says Joel Wertheim, a molecular epidemiologist at the University of California, San Diego, and a co-author of the Zenodo report, referring to GISAID's rules. "It's disheartening."

GISAID reinstated the Zenodo authors' access within 24 hours. A spokesperson told *Nature* that the data hub is now reviewing the situation and has granted authors access while the investigation continues.

The events have sparked a debate in the research community about how and when to disseminate findings based on data found in online repositories — especially when they relate to the highly charged question of the pandemic's origins.

GISAID, a non-profit public–private partnership, is a major disseminator of genomic information relating to viruses. "This is a crucial resource for our work," says Spyros Lytras, an evolutionary virologist at the University of Glasgow, UK, who was not involved in any of the analyses.

With more than 15 million submissions from almost every country in the world, it is the largest database of SARS-CoV-2 genomes "by quite a margin", says Karthik Gangavarapu, a computational biologist at the University of California, Los Angeles, and one of the co-authors briefly locked out of the site. "If you want to analyse COVID-19 from a global perspective, then you do want to use data from GISAID," says Gangavarapu.

Raccoon dogs

The Zenodo report — which has not been peer reviewed — points to raccoon dogs (*Nyctereutes procyonoides*) and other animal species as possible sources of the spillover of SARS-CoV-2 into humans.

The same data analysed in the report had formed the basis of a preprint[2] posted on Research Square in February 2022 by researchers at the Chinese Center for Disease Control and Prevention (China CDC) in Beijing, who generated the data. In March 2023, co-authors of the Zenodo report noticed that the data were publicly available on GISAID and downloaded them, but then the data disappeared from public view following a request from the data generators.

GISAID's rules require researchers to "make best efforts to collaborate" with data generators and involve the generators in their analyses.

In a public statement on 21 March, GISAID said that it had reached out to the the China CDC and established that the authors of the Zenodo report had not fulfilled this requirement and had communicated "*only* their intent to publish the data generator's data".

GISAID also said that the authors of the Zenodo report had "scooped" the China CDC authors by posting their analysis on Zenodo before the China CDC group's work was published in a peer-reviewed journal. A spokesperson for GISAID told *Nature* that the China CDC had submitted a complaint, stating that the Chinese researchers had asked the Zenodo report authors to "hold off" with their publication because they "had a manuscript pending, and were in the middle of updating some of the data they had submitted at the request of the reviewers". (The China CDC preprint is

currently marked as ‘under review’ at Nature Portfolio; *Nature*’s news team is independent of its publisher, Springer Nature, which publishes Nature Portfolio journals.)

The researchers at the China CDC contacted by *Nature* have not responded to requests for comment on the situation.

Members of the international team deny these accusations. In their report and an e-mail sent to GISAID on behalf of all the authors by Michael Worobey, an evolutionary biologist at the University of Arizona, Tucson, they say they made multiple attempts to collaborate with the data generators that were declined. The group acknowledges the China CDC team in the analysis.

Wertheim says that the fact that a preprint has already been posted online protects the China CDC researchers from being scooped. And in their Zenodo report, the authors explicitly state that they respect the China CDC researchers’ right to be the first to publish a peer-reviewed paper on their data and do not intend to submit their report for review, which would compete with that process.

The Zenodo authors also point out in their report that more than three years have passed since the collection of the data, which are critical for the understanding of the pandemic’s origins.

Sparks discussion

The debate over the Zenodo report has sparked discussion in the wider research community. By posting a report in advance of the data owners’ work being published in a journal, the Zenodo researchers could deter data generators from sharing information for fear of being scooped, says Gustavo Palacios, a virologist at the Icahn School of Medicine at Mount Sinai, New York City. This could impair the response to future pandemics, he says.

But the fact that a preprint describing the data has already been posted protects the data providers, say other researchers. The pandemic has set a precedent of using preprints to get credit for data and analysis, says Sergei Pond, a computational evolutionary biologist at Temple University in

Philadelphia, Pennsylvania. “As soon as you put up a preprint, that stakes your claim that these are your data,” he says.

Senjuti Saha, a molecular microbiologist at the Child Health Research Foundation in Dhaka, questions the need for urgency on the part of the Zenodo authors: she says people’s lives don’t depend on the data at this time and therefore there was no need to rush out an analysis.

The data’s relevance to the pandemic’s origins “is not a good enough reason to undermine a fantastically beneficial open-data enterprise that is entirely dependent on both the goodwill of data producers and the good behaviour of data users”, says Darren Martin, a computational biologist at the University of Cape Town, South Africa.

But rather than an attempt to scoop the China CDC work, others see the Zenodo report as the sort of widely accepted and beneficial scientific discussion that should ensue from a preprint. “Other people can start to analyse the data and can write about their interpretations of the data. And they can do that citing the original researchers, giving them the credit for generating those data,” says Theo Sanderson, a geneticist at the Francis Crick Institute in London, who highlights that his views don’t represent those of his institution.

Lack of clarity

Several researchers also say that there is a lack of clarity in GISAID’s data-access rules and how they are enforced, and, given GISAID’s importance for scientific research, clarity is crucial.

The rules don’t explicitly say anything about scooping, don’t provide clear definitions for what entails “best efforts to collaborate”, don’t specify a time limit for when that requirement to request collaboration ends, and don’t say whether a group may publish using GISAID data if the data producers decline or do not respond, say researchers *Nature* spoke to.

Greater clarity from GISAID as to the circumstances in which access to GISAID data could be revoked would prevent researchers from unwittingly violating the rules, says Gangavarapu.

GISAID says it expects users to apply “common sense” and seek clarification directly when they are unsure about the rules. It says it does not explicitly refer to ‘scooping’ in its terms of use due to the phrase’s broad interpretation. And if parties disagree on collaboration, this could result in separate publications “but there always needs to be priority for the data generator to tell their story first”.

GISAID describes the recent situation as an “extremely rare event” and says that suspending an account is a measure of last resort. “The vast majority of GISAID’s tens of thousands of users demonstrate sound scientific etiquette which has contributed to the initiative’s overwhelming success over the course of the last 15 years,” says GISAID.

After the Zenodo report was posted, GISAID made the data at the heart of the analyses available to its users once more. It also introduced more restrictions for access to those data, with a pop-up explicitly stating that any analyses based on the data will not be publicly shared. This pop up has now been removed.

Nature **616**, 13-14 (2023)

doi: https://doi.org/10.1038/d41586-023-00901-9

Updates & Corrections

- **Update 31 March 2023**: GISAID has now removed the additional restrictions that it had introduced for access to the China CDC data; the story has been changed to reflect that.

- **Correction 31 March 2023**: An earlier version of this story said that GISAID had clarified its rules for access to the China CDC data after the Zenodo report was posted; in fact it added more restrictions.

References

1. Crits-Christoph, A. *et al.* Report at Zenodo https://doi.org/10.5281/zenodo.7754299 (2023).

2. Gao, G. *et al.* Preprint at Research Square https://doi.org/10.21203/rs.3.rs-1370392/v1 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00901-9

- NEWS
- 29 March 2023

Researchers in Germany protest proposed postdoc rule change

A potential legal reform is reigniting debate about a time limit on fixed-term postdoc contracts.

- Layal Liverpool



Last week’s protest in Berlin was part of a long-running debate on a law about fixed-term contracts for researchers.Credit: Mike Cowburn

(cowburn86)

Hundreds of researchers gathered in front of the German research ministry in Berlin on 24 March to demand improvements to working conditions in academia. The protest is the latest flashpoint in a long-running debate about a law that means postdocs can be employed on fixed-term contracts for only up to six years after they complete their PhD.



Lab leaders wrestle with paucity of postdocs

“This was the biggest protest we have ever managed to create,” says co-organizer Sandra Janssen, a member of the Network for Decent Labour in Academia, which coordinated the protest. As well as campaigning to abolish the Act on Fixed-Term Employment Contracts in Academia (WissZeitVG), the network is calling for the introduction of permanent contracts for postdocs and for the abolition of habilitation, a thesis-based qualification often required to become a professor.

In Germany, as in many countries, postdocs face tough competition for limited faculty positions. But those who can’t secure permanent jobs before the six-year limit might be forced to leave academia — or the country.

“Germany is losing a lot of great minds to other places,” says Igor Ogashawara, a postdoc in environmental science at the Leibniz Institute of

Freshwater Ecology and Inland Fisheries in Berlin.

Deadline slashed

The protest was part of a backlash against a 17 March proposal by the research ministry to reduce the limit to three years. After two days of widespread criticism from the research community, the ministry said that it was withdrawing the draft bill. It went on to say that it would host a live-streamed open discussion on 30 March to inform the development of a revised draft. “We take this debate very seriously and will once again bring all stakeholders to the table,” a spokesperson for the German Federal Ministry for Education and Research told *Nature.*

The ministry has argued that without a limit on the postdoc phase, it would become more difficult for new generations of researchers to enter academia. It proposed the reduction after consulting stakeholders including universities and research institutions. But academics who spoke to *Nature* said that three years isn’t long enough to find a permanent position.

“Shortening the postdoc period as it was planned is suicidal to the system. No one gets into a competitive position within three years,” says Paula-Irene Villa Braslavsky, a professor in sociology and gender studies at the Ludwig Maximilian University of Munich. “I couldn’t have done it at all.” She started a petition opposing the proposed reform, which has received signatures from almost 2,800 professors at German institutions.



[Postdocs under pressure: 'Can I even do this any more?']

Ogashawara, who is originally from Brazil, says that the postdoc time limit makes things particularly difficult for researchers who aren't German citizens, because applying for a visa and residence permit can be complex and time consuming. "We lose a lot of time, especially non-Europeans," he says. "So, in the end, we do not have a lot of time to do our research."

Janssen, a postdoc in comparative literature and the history of science at the Institute for Advanced Study in the Humanities Essen, says one problem is the lack of permanent positions available. She thinks that eliminating the WissZeitVG entirely would force research institutions to create more permanent positions, instead of continually hiring postdocs on fixed-term contracts with limited opportunities for career progression.

Villa Braslavsky says that the postdoc phase coincides with when many people start having children — as it did for her. The WissZeitVG allows extensions to the six-year limit for researchers who become parents or have a disability or chronic illness. But taking time out remains a disadvantage in terms of career progression, she says.

Brain drain

The time limit and lack of permanent positions also affects Villa Braslavsky as a principal investigator. “It is often difficult to retain excellent people,” she says. In a statement published online in response to the proposed reform, 24 academics who left Germany to pursue research careers in other countries warned of a “brain drain”.

Thomas Kirchner, a physicist at the Martin Luther University of Halle-Wittenberg who co-organized the Berlin protest, is among those considering leaving. “I have family in Germany, I have friends in Germany, I would like to stay in Germany, but if I’m on an open playing field and I don’t have these constraints, then I’m not choosing Germany. I mean, why would I?” he says.

Immediately after the protest, Kirchner says, the Network for Decent Labour in Academia received many requests for communication from the research ministry. “I think they heard us a little,” he says. But he isn’t optimistic about the open discussion on 30 March. “My biggest fear actually is that it will just result in a stalemate, so now nothing changes.”

Nature **616**, 15 (2023)

doi: https://doi.org/10.1038/d41586-023-00913-5

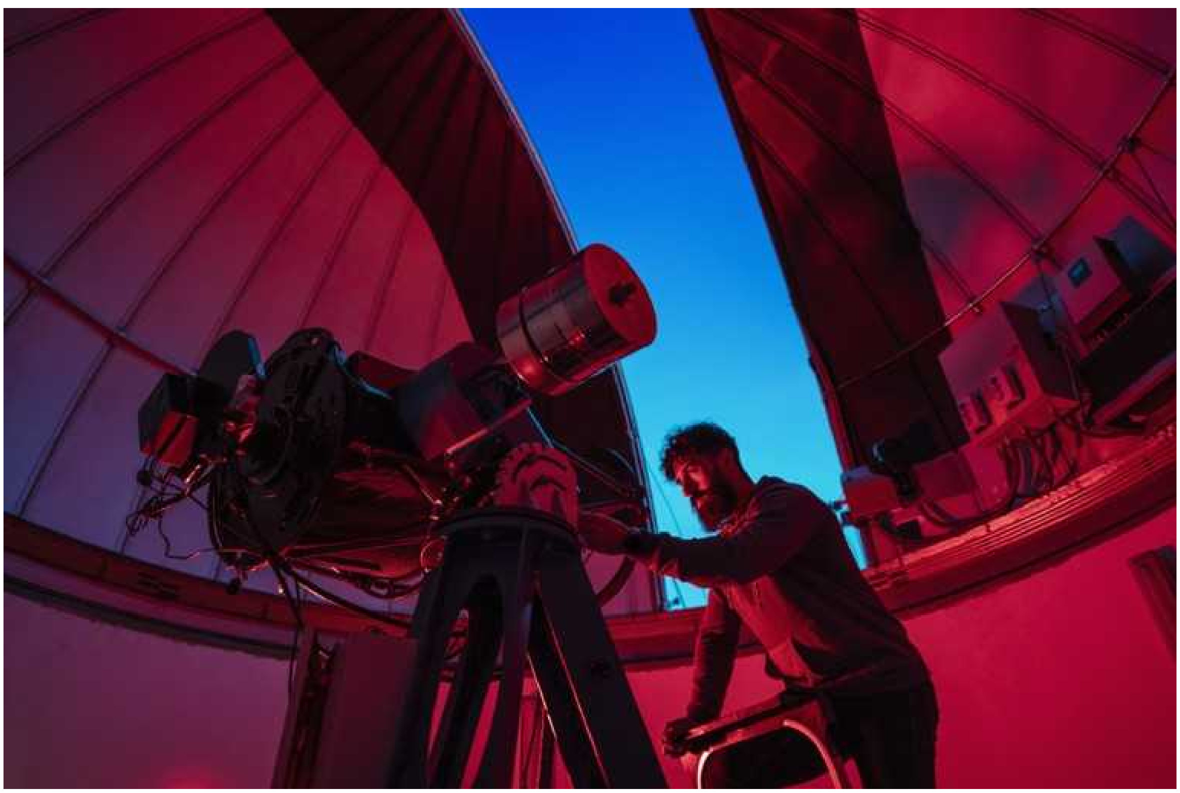
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00913-5

- NEWS
- 28 March 2023

How green is your research? These scientists are cutting their carbon footprints

As the effects of climate change grow, scientists from fields spanning astronomy to biology are trying to decarbonize their research.

- Myriam Vidal Valero



Astronomical observatories require vast amounts of energy to run.Credit: Getty

Michael Inouye, a computational biologist at the Baker Heart and Diabetes Institute in Melbourne, Australia, has always worried about the carbon emissions that his research was generating. But it was the unprecedented wildfires in Australia in 2019 and 2020, which were made more likely by climate change and shut down his laboratory, that pushed him to take action. In 2020, he began collaborating with two PhD students to measure the carbon emissions of their computational research. The project created Green Algorithms, a free online calculator that allows users to estimate the carbon footprint of their research projects, in a bid to ultimately reduce emissions.

Inouye has been using the calculator since it launched, but last year he took things a step further. After calculating the emissions of one of his team's projects[1] on human genetics and diet, his group planted 30 trees to counter its emissions. Inouye knows that offsetting practices are controversial. Critics of the practice argue that if the trees are not properly monitored, there is no guarantee that they will still be alive the following year, for example. "But I think, it's better to do something than nothing," he says.

He is not alone. As the impacts of climate change grow, researchers from fields spanning astronomy to biology have been working to understand and address the sources of their emissions. But solutions don't come without obstacles.

Archival data

In 2019, a group of scientists at the Institute for Research in Astrophysics and Planetology (IRAP) in Toulouse, France, measured their institution's greenhouse-gas emissions. They included electricity, heating, water, air conditioning, professional travel and the energy required for computing. "The observatories were the biggest contributors to our carbon footprint," says Jürgen Knödlseder, an astrophysicist at IRAP who co-authored a paper on the emissions study[2], published in November 2022.

In 2019, the footprint attributed to the use of observatory data was 4,100 tonnes of carbon dioxide emissions, the equivalent of 2,050 petrol cars running all year in the United Kingdom. You need a lot of electricity to run the observatory, says Knödlseder.

The results spurred the scientists to think about how to decarbonize their work when their main source of emissions is the technology they rely on to make breakthroughs.

One option is to pause the collection of new information and do research on archived data instead, says Knödlseder, who has started such projects with his PhD students. “When you think about it, you prepare an observatory for a couple of years, then the thing is launched and you do two or three years of data analysis, and then you jump to the next project. But the orbiting satellite is still running, it is still getting good data,” he says.

Clinical-trial footprints

Biomedical researchers face similar issues. “Clinical trials, especially when designed in certain ways, can be very carbon intensive,” says Rustam Al-Shahi Salman, a neurologist at the University of Edinburgh, UK. In 2021, he co-authored an article[3] with the Sustainable Healthcare Coalition in London — an organization that promotes green practices in the sector — proposing a strategy for measuring the carbon footprint of clinical trials. The tool is being tested and the results should be available later this year, says Al-Shahi Salman.

Previous studies have focused on estimating emissions from buildings and from researchers, patients and staff travelling to the trial site, Al-Shahi Salman says. For example, a 2011 study found that the carbon emissions of two clinical trials that tested the effects of corticosteroids in people with brain injuries ranged from 108.2 to 181.3 tonnes of carbon emissions[4]. The transport of trial materials and the energy to power the trial-coordination building had the largest footprints.

One way to reduce emissions would be to continue using telemedicine strategies implemented during the COVID-19 pandemic, such as holding meetings, gaining patient consent and doing follow-ups by video, or asking patients to wear health-monitoring devices.

But some parts of the clinical-trial process require institutional interventions to make them greener, such as the manufacturing and distribution of drugs

and placebos, says Al-Shahi Salman.

Culture change

Reducing research emissions on a large scale requires institutional support, says Pierrick Martin, an astrophysicist at IRAP. “The situation is not that different from the one you can experience as a citizen. There are things you can do yourself, at your level, in your local environment, that are worth something. But this has limits, and you can’t escape from political decisions at some point.”

And the time to reach a consensus is running out. The Intergovernmental Panel on Climate Change set the net-zero emissions goal for 2050 — a demanding target, says Martin. The closest the world has come to cutting emissions that drastically was during pandemic lockdowns, which cut global CO_2 emissions by close to 10%. “This is what we are supposed to achieve every year,” says Martin.

Institutions should start studying how they can reduce their carbon footprint instead of just measuring it, Martin adds. “If everybody starts to count carbon-footprint assessments, this will take years,” he says. “As soon as you have a few carbon-footprint assessments here and there in different countries or different infrastructures, you start to have a pretty good idea of what you need to do.”

One way to push things forward would be for institutions to ask scientists for emissions-reduction strategies as a condition of grants or jobs. But this is challenging for research institutions who are not used to considering their emissions, says Knud Jahnke, a cosmologist at the Max Planck Institute for Astronomy in Heidelberg, Germany. “It’s fundamentally human and fundamentally challenging for big organizations because if they’re big, they’re well-oiled in the way they’re used to operate.”

Nature **616**, 16-17 (2023)

doi: https://doi.org/10.1038/d41586-023-00837-0

References

1. Qin, Y. *et al. Nature Genet.* **54**, 134–142 (2022).
2. Martin, P. *et al. Nature Astron.* **6**, 1219–1222 (2022).
3. Adshead, F. *et al. Lancet* **398**, 281–282 (2021).
4. Subaiya, S., Hogg, E. & Roberts, I. *Trials* **12**, 31 (2011).

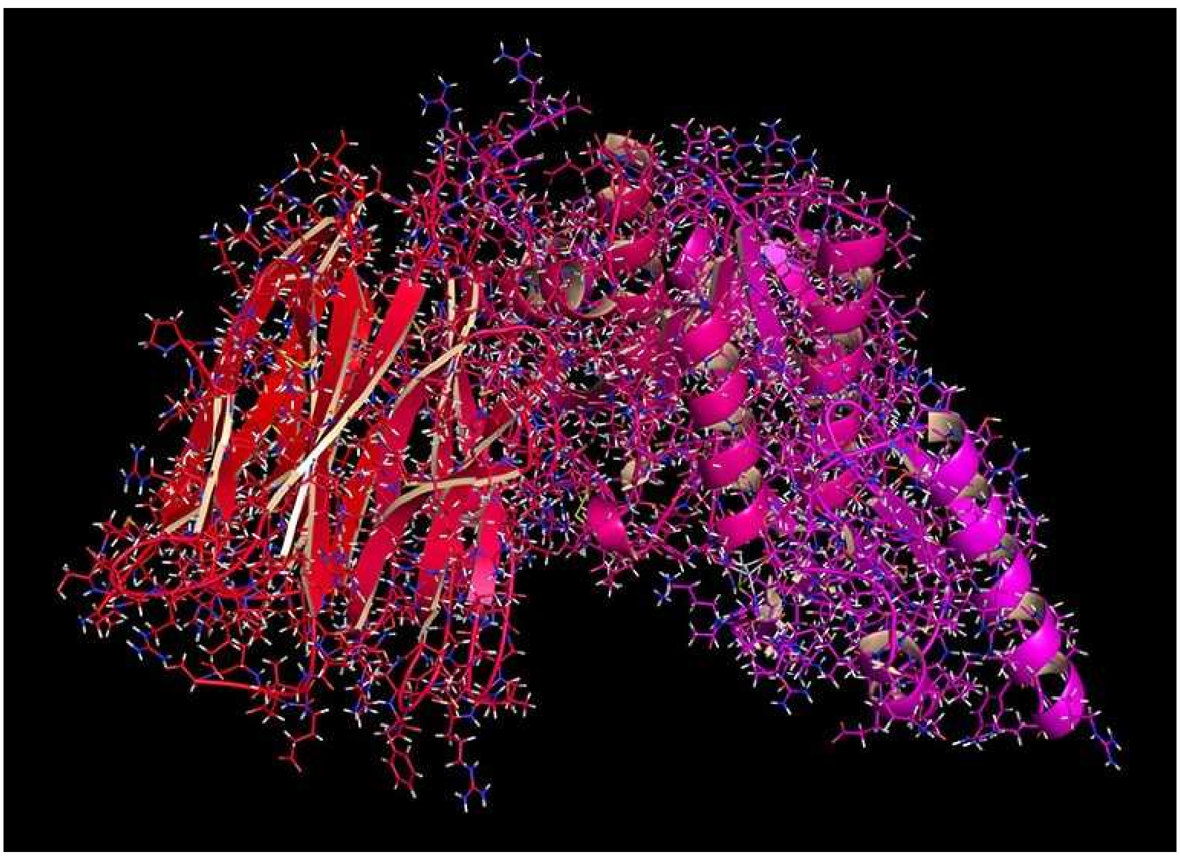
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00837-0

- NEWS
- 28 March 2023

Antibody-patent row could have far-reaching impact on biotech

The results of a US case could dictate how broad patents are, and have knock-on effects for those developing drugs.

- Heidi Ledford



The PCSK9 protein is the target of several antibodies that are used to treat high cholesterol.Credit: molekuul.be/Alamy

An unusual patent case before the US Supreme Court could have wide-ranging impacts on drug prices and technology.

On 27 March, the court heard arguments in a dispute over rights to therapeutic antibodies that are used to treat high cholesterol in people at risk of cardiovascular disease. Although the court's justices spent much of their time wrestling with the details of how these antibodies are isolated, their decision — expected by the end of June — could affect how specific patents of any ilk must be when they describe an invention, and how broad they can be.

"It could spill over into all types of biotechnology cases," says Sean Tu, a legal scholar at West Virginia University in Morgantown. "Today we're talking about antibodies, tomorrow we might be talking about CRISPR or CAR-T-cell therapies."

Cholesterol clearance

At the core of the current case are antibodies that bind to a protein called PCSK9, which decreases the clearance of 'bad' cholesterol from the blood. The antibodies block PCSK9 activity, resulting in lower levels of this cholesterol. The first such antibody, called Praluent (alirocumab), was approved by the US Food and Drug Administration in July 2015 and was developed by two pharmaceutical companies: Sanofi in Paris and Regeneron in Tarrytown, New York.

Since then, several other PCSK9 therapies have come to market, including a rival antibody called Repatha (evolocumab) made by Amgen in Thousand Oaks, California, that was approved in August 2015. Sales of Repatha reached US$1.3 billion last year; sales of Praluent were $467 million.

In the case before the US Supreme Court, Amgen is appealing against a federal court ruling that struck down one of its key patents on Repatha. The patent claims not only a handful of specific antibodies, as defined by their amino-acid sequence, but also any other antibodies, regardless of their structure, that bind to the same region of the PCSK9 protein.

This, said the lower court, was insufficiently detailed. US law requires that a patent contain a written description of the invention that provides enough detail for someone “skilled in the art” to use the patent as a recipe to recreate the invention. But Amgen’s patent, the federal court argued, merely disclosed how to sift through antibodies found in mouse serum for those that bind to a region of the PCSK9 protein, rather than first specifying what the sequence of those antibodies would be. “They have not given the information that a person skilled in the art would need to make and use all of the antibodies,” said Colleen Sinzdak, assistant to the US solicitor-general, during oral arguments before the Supreme Court. “The amino acid sequence is the recipe.”

Discouraging development

A win by Amgen could hinder the development of new medicines by restricting competition and discouraging the development of multiple therapeutic antibodies that bind to the same target, says Gregory Winter, a molecular biologist at the University of Cambridge, UK. The availability of multiple antibodies helps keep the price of such drugs — already among the world’s most expensive medicines — in check, while also providing treatment options for people who develop an immune response against one particular therapeutic antibody.

In 2018, Winter shared the Nobel Prize in Chemistry for his work on therapeutic antibodies. When a colleague showed him the patent at issue in the Supreme Court case, he decided to submit a brief to the court explaining the science underlying such medicines.

In the brief, Winter and his co-authors compare antibodies to an attempt to make calorie-free ice cream. Amgen, they said, made vanilla and chocolate ice cream that it determined, after the fact, to contain no calories. “But rather than patent only those two flavors, Amgen claims to have invented all calorie-free frozen dessert, be it ice cream, sorbet, sherbet, or gelato,” they wrote.

Amgen’s attorneys, however, have argued that striking down its patent will harm, not help, future development of therapeutic antibodies. Broad patent

protection is necessary to justify the company's hefty investment in a specific medicine, said Jeffrey Lamken, an attorney who represented Amgen before the Supreme Court. "We've reached an endpoint where, frankly, the industry can't take it any longer," he said. "You can't invest $2.6 billion if the breadth of your claims is such that it means you can't get adequate protection."

Patents 'squeezed'

A decision against Amgen could discourage investors from taking a risk on other biotechnology companies as well, said Michael Penn, vice-president of intellectual property at Instil Bio, a company that is developing cell therapies in Dallas, Texas, at a meeting hosted by the American University Washington College of Law in Washington DC after oral arguments. "When they see patents getting squeezed and patent breadth getting squeezed year for year, that investment goes elsewhere," he said. "That's the fear."

The debate divides the industry into those that focus on discovering drug targets, and favour broad patent protection for their findings, and those that develop specific molecules to drug the targets, says Ulrich Storz, a patent attorney at the Michalski Hütterman and Partner law firm in Düsseldorf, Germany. Universities are more likely to patent drug targets than actual candidate drugs, and the Association of University Technology Managers in Washington DC has filed a brief in support of Amgen.

Even so, there is little evidence that the prevailing practice — which over the past few years has favoured relatively narrow patent claims on antibodies — needs to change to allow such broad patents, says Tu, especially given the potential effect on drug prices and treatment options for patients. "The system that we have right now is working pretty well," he says. "Everybody's getting paid."

Nature **616**, 17 (2023)

doi: https://doi.org/10.1038/d41586-023-00911-7

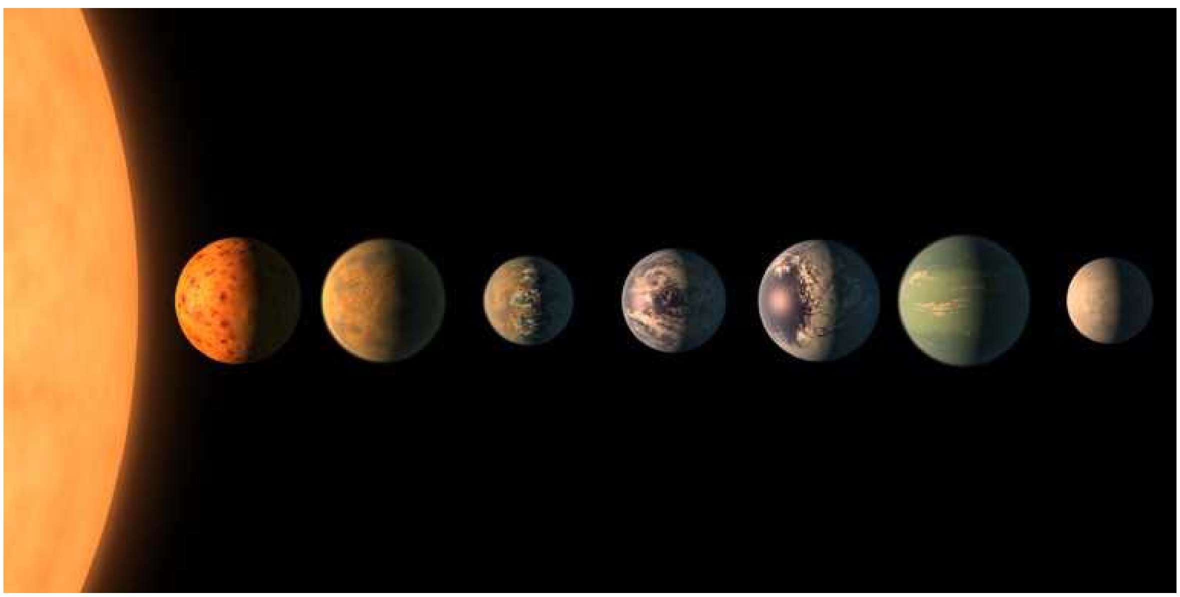
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00911-7

- NEWS
- 27 March 2023

JWST gets best view yet of planet in hotly pursued star system

Telescope didn’t spot an atmosphere on TRAPPIST-1b, but has kicked off a new era in understanding planetary evolution.

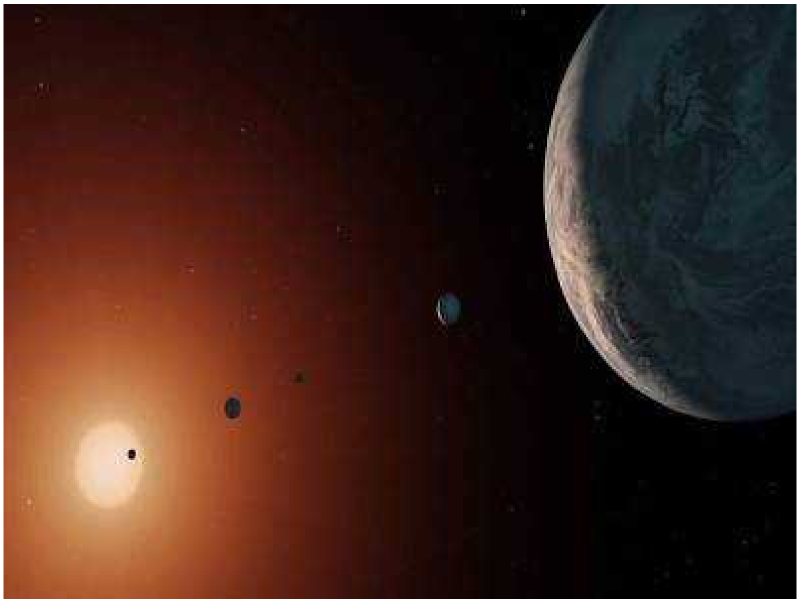
- Alexandra Witze



An artist’s rendering shows approximations of the seven planets of the TRAPPIST-1 system. TRAPPIST-1b (far left) is closest to its star.Credit: NASA/JPL-Caltech

Using the James Webb Space Telescope (JWST), astronomers have confirmed that the planet TRAPPIST-1b probably has no atmosphere. Researchers have been excited to use the new telescope to explore it and its six siblings, which are all roughly the size of Earth and which orbit a star 12 parsecs (39 light years) from Earth. The system is a unique laboratory for

studying how environmental conditions arise on planets — and how they might become suitable for life.



JWST gets first glimpse of 7-planet system with potentially habitable worlds

Last November and December, JWST searched for an atmosphere on TRAPPIST-1b by looking for heat radiating from it. With the telescope's ability to study infrared light, "you can actually measure the glow of the planet", says Thomas Greene, an astronomer at NASA's Ames Research Center in Mountain View, California. He and his colleagues report their results today in *Nature*[1].

Although the finding might sound disappointing to those hoping for an atmosphere, scientists say that the work showcases JWST's transformational power and opens the door for more results from the TRAPPIST-1 system.

Previous studies with the Hubble and Spitzer space telescopes, using a different technique, showed that TRAPPIST-1b — the innermost planet in the system — probably doesn't have a large puffy atmosphere made mostly of hydrogen[2]. But researchers couldn't rule out whether it has a dense atmosphere, as Earth might have had billions of years ago.

A planetary laboratory

JWST looked at TRAPPIST-1 in mid-infrared wavelengths of light — 20 times redder than the human eye can see — to see how that radiation changed as TRAPPIST-1b moved behind the star. By measuring the brightness of the star and planet together compared with that of the star alone, astronomers could calculate how much came from the planet.



JWST reveals first evidence of an exoplanet's surprising chemistry

If TRAPPIST-1b had an atmosphere, it would have re-circulated the energy absorbed from the star and appeared less bright than Greene and his colleagues measured. The observations did not reveal any carbon dioxide on the planet, which JWST might have been able to spot.

It's not surprising that TRAPPIST-1b has no atmosphere, because it is blasted by four times as much radiation as Earth receives from the Sun. TRAPPIST-1 is also wracked by stellar flares and other activity that sends radiation across its planets, potentially scouring away atmospheres. Still, understanding these conditions is crucial because M dwarf stars — cool, dim stars such as TRAPPIST-1 — often have Earth-sized planets orbiting them.

A key early step in studying the TRAPPIST-1 system is to take advantage of JWST's power to understand the star itself, says Julien de Wit, an exoplanet researcher at the Massachusetts Institute of Technology in Cambridge: "know thy star" should come before "know thy planets". Astronomers should work together to study TRAPPIST-1 with JWST using multiple techniques, to understand the star as fully as possible, he says. Otherwise, researchers will struggle to interpret what they are seeing in planetary

observations, because the star's activity could be contaminating those measurements.

More discoveries are sure to come. Other research teams have been using JWST to study TRAPPIST-1b, as well as other planets in the system. That includes TRAPPIST-1b's neighbour, TRAPPIST-1c, a planet that is close enough to its star for JWST to study its glow. Publications on all of these are expected soon.

Nature **616**, 18 (2023)

doi: https://doi.org/10.1038/d41586-023-00876-7

References

1. Greene, T. P. *et al. Nature* https://doi.org/10.1038/s41586-023-05951-7 (2023).

2. De Wit, J. *et al. Nature* **537**, 69–72 (2016).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00876-7

- NEWS
- 29 March 2023

‘Astonishing’ molecular syringe ferries proteins into human cells

Technique borrowed from nature, and honed using artificial intelligence, could spur the development of better drug-delivery systems.

- Heidi Ledford



The bacterium *Photorhabdus asymbiotica* uses molecular spikes to pierce a hole in the membranes of host cells.Credit: F. Zhang *et al./Nature*

Researchers have hijacked a molecular ‘syringe’ that some viruses and bacteria use to infect their hosts, and put it to work delivering potentially therapeutic proteins into human cells grown in the laboratory.

“It’s astonishing,” says Feng Jiang, a microbiologist at the Chinese Academy of Medical Sciences Institute of Pathogen Biology in Beijing. “It is a huge breakthrough.”

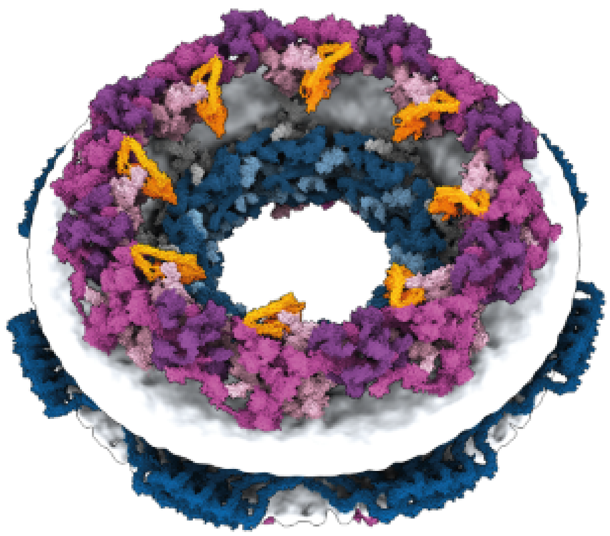
The technique, published in *Nature* on 29 March[1], could offer a new way to administer protein-based drugs, but will need more testing before it can be used in people. With further optimization, the approach might also be useful for delivering the components needed for CRISPR–Cas9 genome editing.

Difficult delivery

The medical applications of CRISPR are currently limited by the challenges of getting the reagents — the DNA-cutting Cas9 enzyme and a short piece of RNA that guides Cas9 to a specific region in the genome — into cells.

“One of the major bottlenecks for gene editing is delivery,” says study co-author Feng Zhang, a molecular biologist at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, and an early pioneer of the CRISPR–Cas9 technique. Limited options have restricted most clinical trials to editing genomes in liver, eye or blood cells, because those cells can be reached using the current delivery methods, he says. “The reason we don’t see brain or kidney diseases getting tackled is because we don’t have good delivery systems.”

While Zhang and his collaborators searched for ways of transporting proteins into human cells, microbiologists were learning more about an unusual group of bacteria that use molecular spikes to pierce a hole in the membranes of host cells. The bacteria then transport proteins through the perforation and into the cell, exploiting the host’s physiology in their favour.



[What's next for AlphaFold and the AI protein-folding revolution]

Last year, Jiang and his colleagues reported that they could manipulate this syringe-like system in the bioluminescent bacterium *Photorhabdus asymbiotica*, loading proteins of their choosing from mammals, plants and fungi into the syringe[2]. Normally, the bacterium lives inside nematodes and uses its syringe to transport a toxin into the cells of insects infected by the nematode. The toxin kills the insect, and the nematode eats the remains. “The bacterium can be viewed as a hired gun to kill this insect,” says co-author Joseph Kreitz, a molecular biologist at the Massachusetts Institute of Technology in Cambridge.

In Zhang’s lab, Kreitz and his collaborators were working on ways to engineer the *P. asymbiotica* molecular syringe so that it would recognize human cells. They focused on a region of the syringe called the tail fibre, which normally binds to a protein found on insect cells. Using the artificial-intelligence program AlphaFold, which predicts protein structures, the team designed ways to modify the tail fibre so that it would recognize mouse and human cells instead. “Once we had the image, it was very easy to modify it for our uses,” says Kreitz. “That was the moment when it all came together.”

The researchers then loaded the syringes with various proteins, including Cas9 and toxins that could be used to kill cancer cells, and delivered them into human cells grown in the lab, and into the brains of mice.

Flexible system

The system was unable to transport the mRNA guide needed for CRISPR–Cas9 genome editing, but the team is developing ways to do this, says Kreitz. The fact that the system was able to ferry Cas9 into cells speaks to the technique's flexibility, he adds, given that the Cas9 protein is about five times larger than the syringes' usual cargo.

The syringe story is reminiscent of the way that researchers such as Zhang developed CRISPR–Cas9 — a system that many microorganisms rely on in nature to defend against viruses and other pathogens — for use as a genome-editing technique, says Asaf Levy, a computational microbiologist at the Hebrew University of Jerusalem. Similar to the early days of CRISPR–Cas9 research, the bacterial syringes are studied by only a handful of labs, and their roles in microbial ecology are only beginning to be understood.

Yet they could have a transformational effect on medicine, says Levy. "The evolution of this thing is quite amazing," he says. "The fact that you can engineer both the payload and the specificity is ultracool."

Nature **616**, 18-19 (2023)

doi: https://doi.org/10.1038/d41586-023-00922-4

References

1. Kreitz, J. *et al. Nature* https://doi.org/10.1038/s41586-023-05870-7 (2023).
2. Jiang, F. *et al. Sci. Adv.* **8**, eabm2343 (2022).

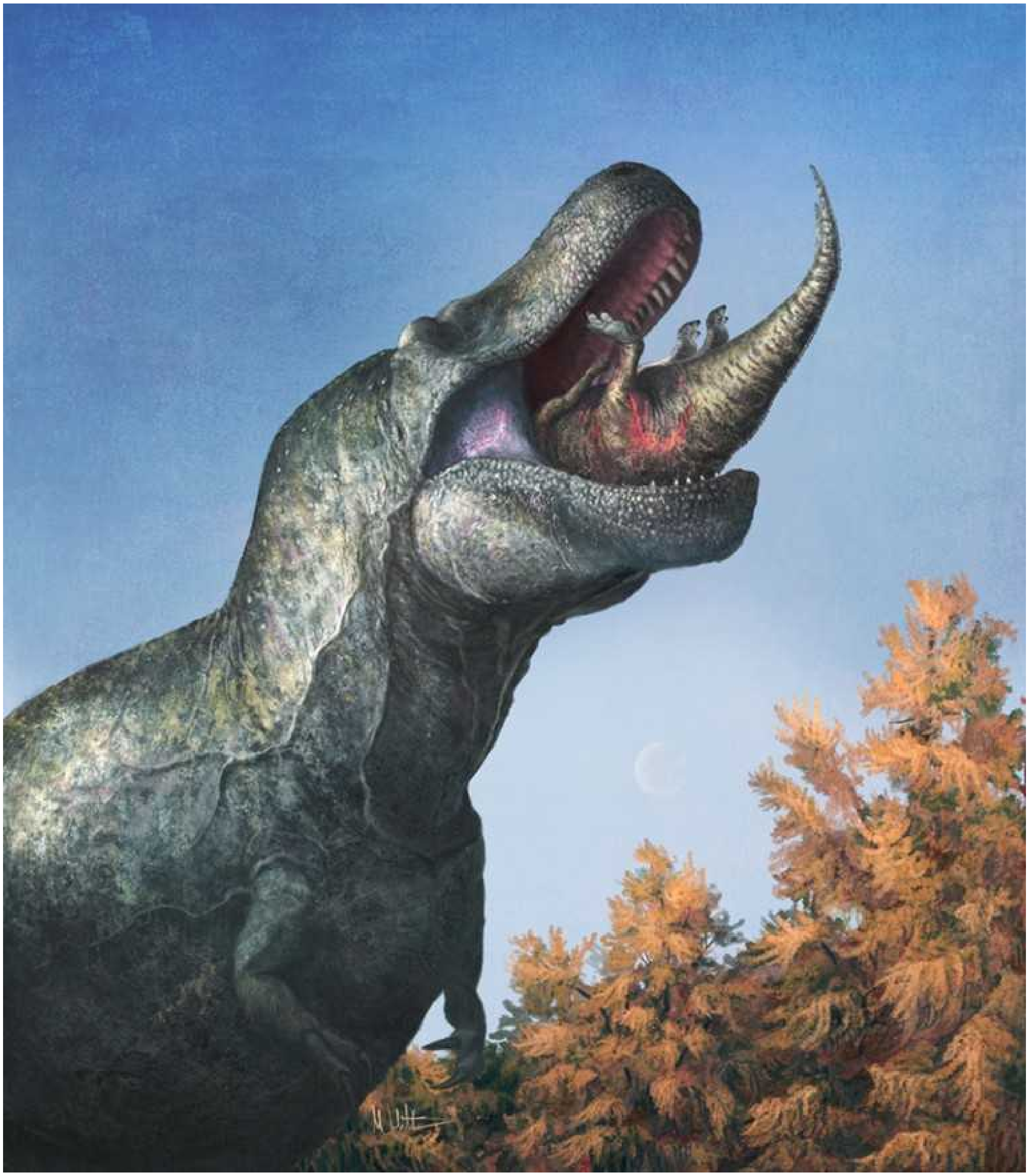
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00922-4

- NEWS
- 30 March 2023

Facelift for *T. rex*: analysis suggests teeth were covered by thin lips

Crocodiles and Komodo dragons provide evidence to support the idea of a scaly cover over the teeth of dinosaur *Tyrannosaurus rex*.

- Dyani Lewis



A juvenile *Edmontosaurus* disappears into the jaws of *Tyrannosaurus*. The reconstruction of soft tissues around the jaws of *T. rex* strongly suggests that it resembled lizards with lips rather than lipless crocodilians.Credit: Mark P. Witton

The fearsome maw of the iconic *Tyrannosaurus rex* has had a makeover. According to a study published today in *Science*[1], the dagger-like teeth of

theropod dinosaurs such as *T. rex* would not have been visible when their mouths were closed. Instead, they would have been concealed behind thin, scaly lips.

Reconstructions of large theropods — whether in Hollywood films or scientific texts — often depict the prehistoric predators as having jaws full of exposed teeth, even when their mouths are shut. Palaeontologists had reasoned that theropod teeth were probably exposed because the teeth are so big, and because the closest living relatives of theropods, alligators and crocodiles, have toothy grins.

But when Kirstin Brink was a graduate student at the University of Toronto, Canada, in the 2010s, she and fellow students Thomas Cullen and Derek Larson got talking over a few beers about whether that depiction was accurate. “It just looks so weird when the *T. rex* mouth is closed, and you can see the teeth,” she says.

Brink, now a vertebrate palaeontologist at the University of Manitoba in Winnipeg, Canada, notes that modern-day lizards, such as iguanas and Komodo dragons, have lips that hide their teeth.

Other palaeontologists have also wondered whether extinct theropods smiled like a crocodile or had lips like lizards. But Brink, Cullen, Larson and their colleagues are the first team to examine the skulls and teeth of theropods and their living relatives to settle the debate.

Brink says that the tooth enamel in theropods points to them having lips that cover the teeth. “Enamel needs to stay hydrated,” she says, otherwise it is prone to cracking.

In crocodiles, the enamel is thick and stays hydrated because they live in the water. Even so, crocodile teeth bear the signs of cracks and damage on their outer surface. That’s not the case in theropods, she says. Theropod teeth are covered by just a thin layer of enamel, indicating that these dinosaurs probably had lips to keep the teeth protected and coated in saliva when their mouths were closed.

Gummy dragons

Brink and her colleagues also compared theropods with Komodo dragons (*Varanus komodoensis*), one of the few living reptiles with teeth shaped like those of theropods. “Komodo dragons are weird though, because they have these gigantic gums,” says Brink. There’s no evidence that theropods had such gummy mouths, but it’s possible, she says.

“These authors have given *T. rex* a facelift,” says Steve Brusatte, a vertebrate palaeontologist at the University of Edinburgh, UK. The work is “the single strongest case for lippy tyrannosaurs yet”, he says.

Soichiro Kawabe, a vertebrate palaeontologist at Fukui Prefectural University in Eiheiji, Japan, says he has long suspected that theropod enamel was too thin for the teeth to be exposed. “So I am happy that the scientific data has now been presented,” he says. “It looks less monstrous and more natural,” he adds. Kawabe would like to see similar work to be done in other prehistoric creatures, such as toothy pterosaurs, for which the facial morphology is less known than for the well-studied *T. rex*.

T. rex wouldn’t have been able to purse its lips like humans do because it wouldn’t have had the necessary muscles. “Dinosaur lips would have been scales and keratin, and some soft bits covering the teeth from the outside, not the big fleshy and pouty lips of humans,” says Brusatte.

“I doubt it will be the final word in this contentious debate,” he says. “Ultimately, we need to find a fossil *T. rex* mummy with skin and muscles and scales still preserved on the head,” he says. “That would tell us once and for all whether there were lippy tissues covering the teeth, or whether the teeth were more exposed.”

Nature **616**, 19 (2023)

doi: https://doi.org/10.1038/d41586-023-00928-y

References

1. Cullen, T. M. *et al. Science* **379**, 1348–1352 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00928-y

- NEWS Q&A
- 30 March 2023

How to convince a politician: a science adviser’s lessons from the pandemic

Patrick Vallance, who advised the UK government during the height of COVID, says that scientific officials should know four key things when guiding policymakers.

- Katharine Sanderson



UK government chief scientific adviser Patrick Vallance gave many televised briefings during the COVID-19 pandemic.Credit: Leon Neal/Getty

Patrick Vallance has come to the end of his five-year tenure as chief scientific adviser to the UK government, during which he oversaw the unprecedented science response to the COVID-19 pandemic.

Before coming to government, Vallance spent most of his career in academia as a medical researcher focusing on vascular biology, and 11 years in industry, including as head of research and development at the UK pharmaceutical giant GSK.

During the pandemic, he was thrust into UK living rooms to give daily televised updates to the nation, and was charged with providing scientific advice that would guide the country's COVID-19 policy. One of his next roles will be as chair of the board of directors of the Natural History Museum in London.

Nature caught up with Vallance to reflect on his time in office.

When you were appointed in 2017, Brexit was going to be the biggest issue, but it didn't work out that way. How have the past few years been?

Two-and-a-half years were completely taken up with COVID. We did lots of other things but that dominated my day-to-day activities.

Before I came in, I said that I wanted to make sure that we got science and engineering as embedded in the government as we possibly could. Each department needs a strong, externally appointed chief scientific adviser, because that individual is the link between the inside and the outside.

If you want the government chief scientific adviser to be influential, you can't do it on your own. So the fact that we've got an effective network of chief scientific advisers across every department is really important.

How was your COVID-19 experience?

It was all-consuming and it showed the importance of having a strong science base in the United Kingdom. My job was made easier by the fact that I could call on experts from academia and industry. We could get people around the table quickly. That was crucial during things like setting up the vaccines task force.

In areas where it was possible to move quickly, such as genotyping of the virus, where the United Kingdom got off to a good start, it was in part because the infrastructure already existed. It was possible to get vaccines moving because we had good academic research, but also we had big companies that knew how to turn things into industrial products. We were less well served on the scaling up of testing because we didn't have a big diagnostics industry.

Do you think that the response to COVID-19 will have a knock-on effect across medicine?

If you look at what happened with the trials during COVID, the RECOVERY trial was fantastic because it was big enough to get proper answers. Around the world there were thousands of studies that were too small. That's a wasted resource. At the peak of the pandemic, something like 12% of patients in UK hospitals with COVID were in the RECOVERY study — that's remarkable. There's definitely a lesson about how the acceleration of getting answers can be embedded in everyday practice.

What was the United Kingdom's worst pandemic moment, and its best?

When you see people going into hospital, hospitals overflowing, hear the stories of the tragedy that occurred, that can't feel anything other than absolutely awful.

The highlights were the moment we knew something worked, like the drug dexamethasone from the RECOVERY trial, and the moment we got the first readout from the vaccine study that showed it had more than 90% efficacy. That was much, much more than people had even dared hoped for.

The other lesson I've learned is that, often, the louder the voices, the lower the evidence base. During COVID the things that became most contentious and where people had extremely strong views and policy opinions were the ones where the evidence base was weakest. You can see that in the initial uncertainty around facemasks and other areas.

Did politicians listen to your advice?

There are four things that a science adviser needs to ask themselves. First, is the evidence base adequate and, if not, what needs to be done? Second, has the advice been understood? That's really key. It's not only about whether you delivered it. And you need to think about how you would assure yourself that it had been understood.

Point three is: has the science advice being given in a form that is relevant to policy consideration, and what do policymakers need to understand to formulate policy? It's not easy.

The fourth area is to make sure that the role of science in monitoring the policy impact has also been thought through. That's obviously crucial when you think about things such as the climate. When you make policy decisions, how are you going to know you're actually headed in the right direction?

What does Brexit mean for UK science?

I'll be absolutely clear about this: science is global. It's international — we need to have extremely strong links to other countries. We have historically had those links. And it's important that we not only maintain those, but also continue to grow them. Many countries are now doing science and engineering at a high level.

I've been very clear that I believe that the United Kingdom being part of the European Union's Horizon Europe programme is important. It's a system that's taken a long time to get running. It runs very well. It provides good links across Europe, at a scale that is not possible on your own.

What has it been like working with politicians?

The biggest thing coming into government and working with politicians is to understand exactly what it is that enables politicians to listen to science advice, because it's usually not at the top of their mind. We need to recognize that although we find it fascinating, not everyone does. And, therefore, you need to win people over.

An early briefing on climate with three or four of us in the room and the prime minister at the time, Boris Johnson, was instrumental in him changing his mind about what he wanted to prioritize. He wanted to prioritize climate after that. It wasn't because we were campaigning, it was because we were neutrally showing him factual graphs about what was going on and helping to make sure he understood the data. That's what we need to do, not getting dragged into political or policy opinions.

And if I've got one piece of advice for any science adviser, it's: stick to the science.

If after a meeting like that the prime minister is not persuaded, have you failed as an adviser?

You failed as an adviser if they haven't understood it. If they really understood it, and chose to make a different decision, that seems to me that that is the democratic process. I fundamentally and strongly believe in democracy.

Nature **616**, 20 (2023)

doi: https://doi.org/10.1038/d41586-023-00938-w

This interview has been edited for length and clarity.

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00938-w

- NEWS FEATURE
- 05 April 2023

Diversity in German science: researchers push for missing ethnicity data

The European country is one of several reassessing its cultural unease with collecting information on scientists’ race and ethnicity.

- Hristio Boytchev

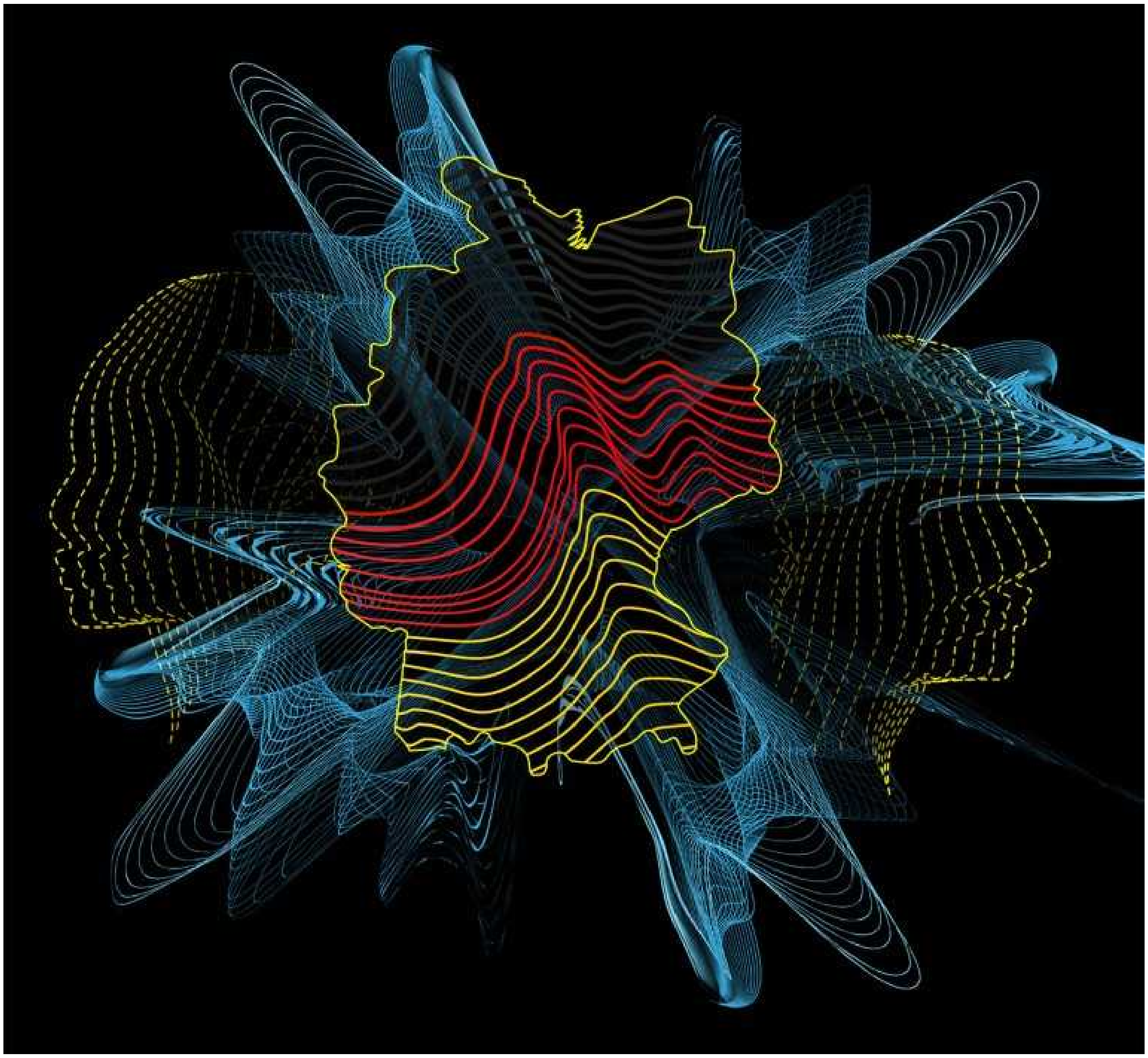


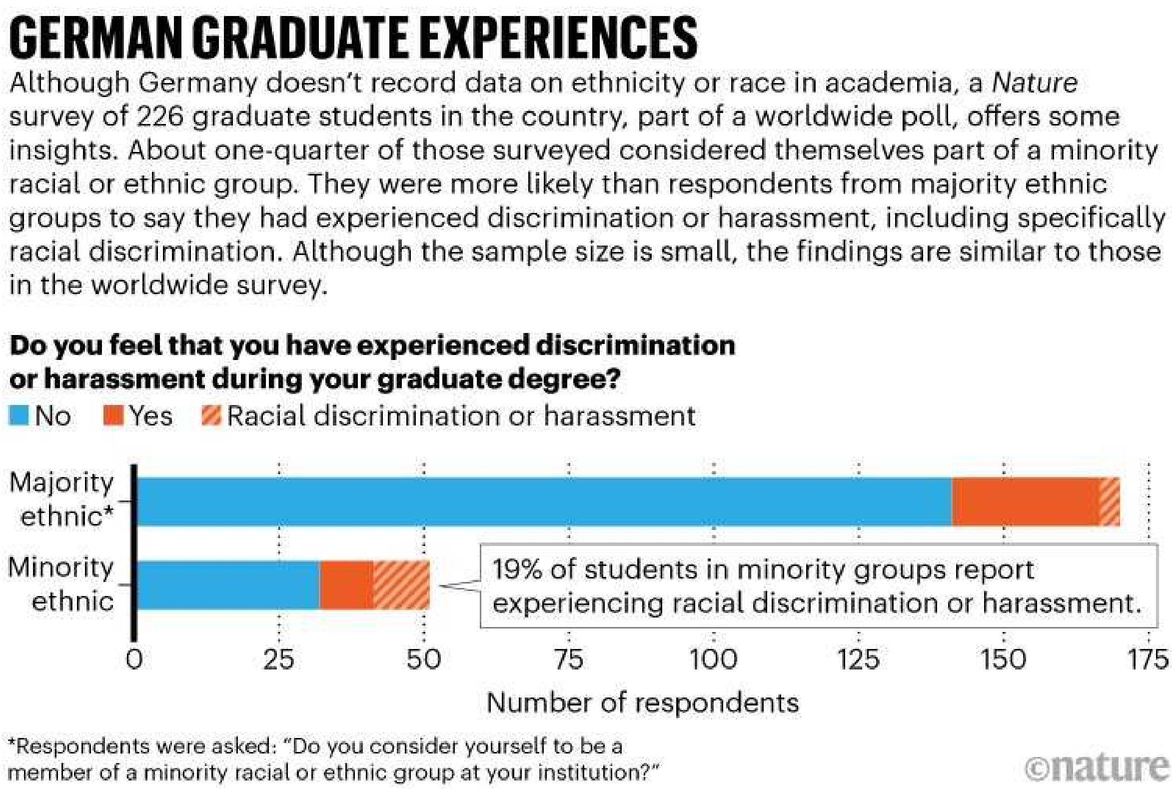
Illustration by Polygon8

This article is part of a Nature *series examining data on ethnic or racial diversity in science in different countries*. *See also: How UK science is failing Black researchers and How India's caste system limits diversity in science*.

In November 2021, Isolde Karle's university in Germany appointed her to a newly created high-ranking role, which signalled the attention it was paying to diversity. As vice-rector for diversity, inclusion and talent development at the Ruhr University Bochum, Karle's job involves helping students from minority groups — including those discriminated against because of their skin colour — to participate fully in science.

“There is racism in society, and we universities are not free of it,” says Karle, who is also a theologian and the university’s pastor.

Survey data support Karle’s concerns. This January, for instance, Germany’s government issued its first annual report on racism. It cites, among other data, a large telephone survey that found more than one-fifth of people in the country had personally experienced racism (see go.nature.com/3zqsmvt; in German). And at universities, a *Nature* worldwide survey of graduate students last year found that, in Germany, 6% had experienced racial discrimination or harassment, but that this rose to 19% among those who identified as being in a racial or ethnic minority; findings similar to the survey’s global results (see ‘German graduate experiences’).



Source: *Nature* graduate survey 2022 (*Nature* https://doi.org/j4cn; 2022)

Karle has talked to some researchers from minority ethnic groups at her university about their experiences. But she can’t clearly track racial or ethnic diversity among the scholars at her institution or how it might be changing, because she has no data. That is the norm for employers all over Germany,

where there is general unease at the idea of collecting information on ethnicity — even though doing so is not legally prohibited — as well as a feeling that it is not a priority issue, according to many *Nature* talked to.

Such concerns are often explained with reference to the country's history and the horrific example of how the Nazis used census data to organize genocide. As a result, sensitivity over data protection and privacy is high: surveying people in Germany about their ethnicity has a strong negative connotation.

Karle expresses these reservations, too; asking employees about their race or heritage can itself be discrimination, she explains. "We want to make the differences that lead to discrimination disappear, but by asking about them, we would focus on them," she says. "This is part of the paradox."

The result in Germany is that, unlike in the United Kingdom, United States and several other countries, there are no census data on racial or ethnic diversity, and neither research funders nor universities collect this information. A spokesperson for the Bonn-based German Rectors' Conference, an association of 269 universities in Germany, told *Nature* that there was no recording of personal characteristics that could identify minority ethnic groups for "understandable" and "historical" reasons, and because of data protection.

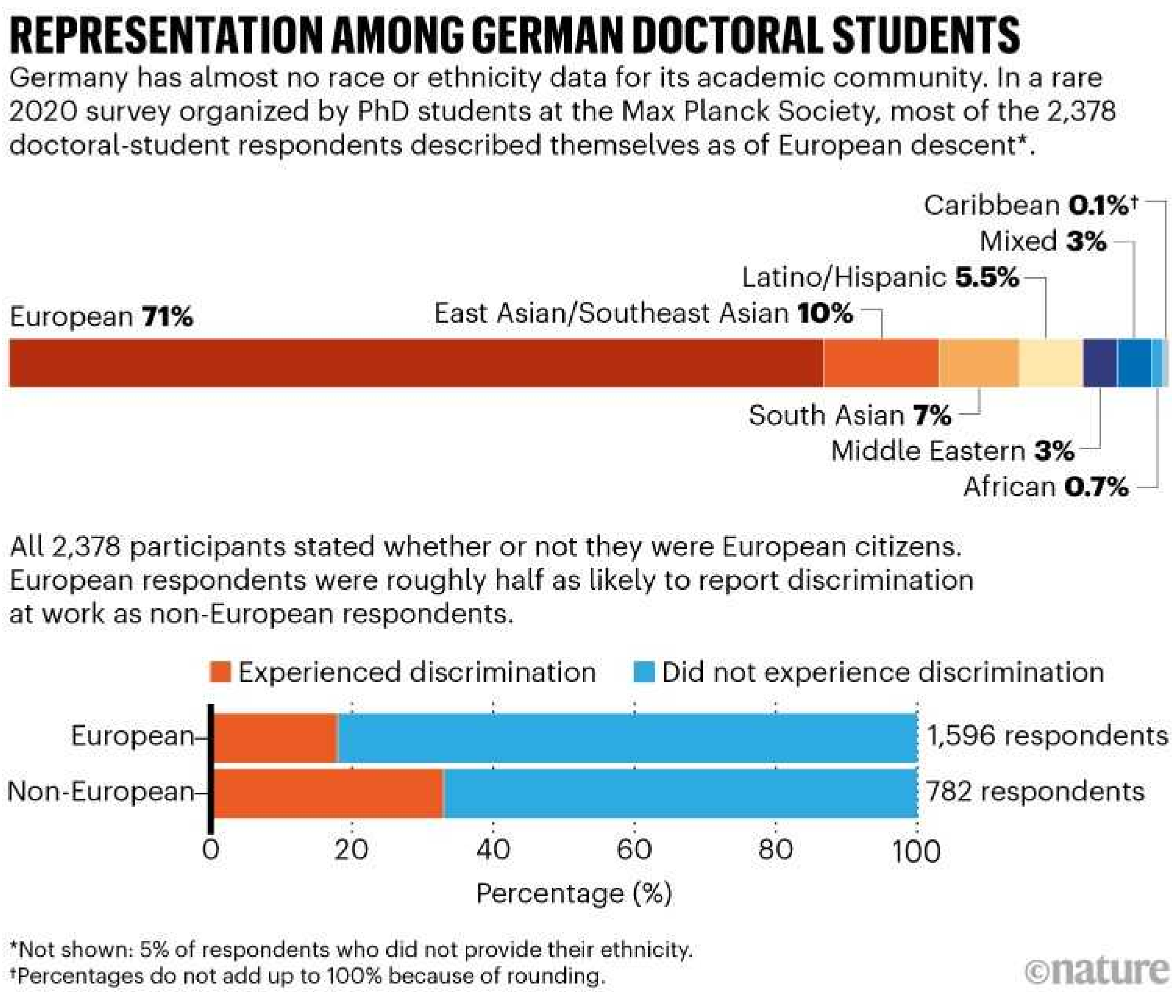
It is a similar situation in some other European countries: three research-funding agencies in France, Sweden and the Netherlands, for instance, told *Nature* that they don't collect diversity data on race or ethnicity, generally citing data-protection concerns.

"Of course, this makes it difficult to quantitatively measure success on ethnic diversity issues," says a spokesperson for Germany's Max Planck Society (MPS) in Munich, a large public association of research institutes.

But this cultural unease is starting to shift. Some researchers in countries including Germany are pushing to gather data on ethnicity and race in academia, arguing that the advantages of quantitative data collection outweigh the concerns. In July 2022, Germany's largest public research

funder, the DFG, issued guidelines that put a greater emphasis on ethnic diversity in data gathering.

For now, there are only snapshots of information. In 2020, an association of PhD students at the MPS, called Max Planck PhDnet, organized its own survey, asking more than 2,000 doctoral students about their citizenship and ethnicity (P.-G Majev *et al. 2020 Survey Report* https://doi.org/j39n; Max Planck PhDnet, 2021). It reported that 71% of the students who provided an ethnicity described themselves as of European descent, 10% as of East Asian or Southeast Asian descent, 7% as of South Asian descent and 0.7% (16 people) as of African descent. One-third of people describing themselves as non-European said they had felt discriminated against at work, compared with 18% of respondents identifying as European (see ‘Representation among German doctoral students’).

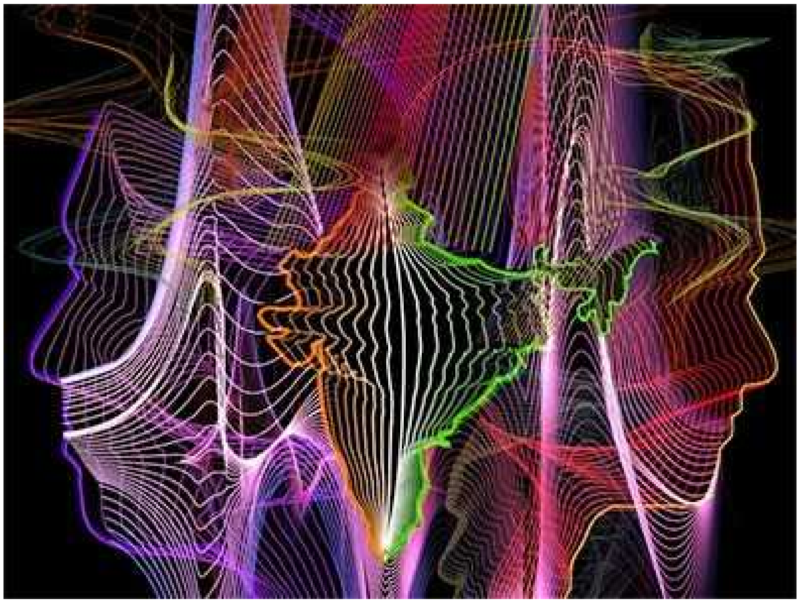


Source: P.-G Majev *et al*. https://doi.org/j39n (Max Planck PhDnet, 2021)

These findings are comparable to those in *Nature*'s 2022 survey: almost one-quarter of German respondents identified as being in an ethnic or racial minority and, of those, more than one-third said they had experienced discrimination or harassment (not always racial); compared with 17% of the other German respondents.

Larger surveys have tended to collect data using a coarse proxy called migration background. This typically refers to a person with at least one parent who was not born a German citizen — which doesn't necessarily map well on to minority racial or ethnic status. In 2019, the German Centre for Higher Education Research and Science Studies (DZHW), a research institute in Hanover, launched a longitudinal study called the National Academics Panel Study that aims to examine the demographics and career paths of doctoral researchers in Germany. It has started to collect data on factors that include migration background and country of birth, but not ethnicity (see go.nature.com/3kt68jv; in German).

The DZHW's publicly available data define migration background as people who were born outside Germany. It found that this group accounted for 24% of some 14,000 doctoral researchers surveyed in 2019–20. (Data using the wider definition of migration background — that is, including PhD students who have a parent born outside Germany — are available to researchers on request, the DZHW says). In a separate study of around 18,000 German students (including undergraduates) last year, the DZHW found that 6% of all students said they had been discriminated against because of their migration background (see go.nature.com/3kgcdj7; in German).



How India’s caste system limits diversity in science — in six charts

And last September, the German Education Union (GEW), which represents researchers and teachers, analysed data from multiple sources to suggest that 2% of German schoolchildren with a migration background might go on to gain a doctorate, whereas those with two German-born parents had double the chance, at 4% (see go.nature.com/3tzwcuj; in German). (The union added that the main differences occurred in transitions from primary school to high school, and from high school to bachelor’s degrees, rather than later in the academic pipeline.)

Such reports leave many questions about diversity in German science unanswered. But the picture might be changing. The Berlin non-governmental organization Citizens for Europe (CFE), for instance, aims to survey the full diversity of people in Germany — including the interplay of race, culture and social class, as well as gender. In 2021, CFE and another Berlin-based non-governmental organization, Each One Teach One (EOTO), released the results of ‘Afrozensus’, a survey of more than 6,000 “Black, African and Afro-Diasporic” people in Germany, which aimed to raise awareness about the experiences of Black people in the country (see go.nature.com/3nptj9m; in German). It did not specifically analyse academic professions, but more than 4% of respondents had a PhD.

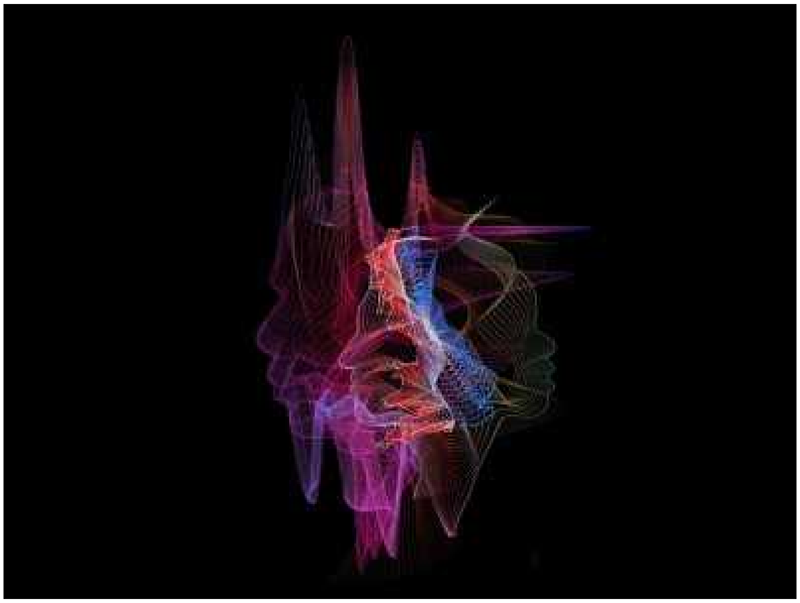
In all, only 15.3% of Afrozensus respondents said they had never been discriminated against in a professional context. The lack of data on race and discrimination is part of a vicious circle, says Daniel Gyamerah, who worked on the survey while at CFE. He is now head of a new non-profit think tank, called zedela (Centre for Data Driven Empowerment, Leadership and Advocacy), and is also EOTO's chairperson. Without data, it is hard to grasp the extent of discrimination problems, Gyamerah says, which then makes it harder to convince organizations and policymakers that they should track the issue. Reservations over data-protection laws are often used as an excuse to avoid engaging with the topic, he says.

Sociologist Lucienne Wagner, a CFE researcher who is pursuing a PhD at the University of Koblenz-Landau on the discourses of diversity at German universities, sees a further challenge for raising awareness about racial discrimination and collecting data on diversity at universities. With scant resources, people focused on gender equality can end up competing against — rather than working with — those exploring other dimensions of discrimination. But efforts in the field mean that attitudes in academia are changing, she adds.

The DFG's diversity guidelines, issued in July, could mark a significant shift. "Transparency of equality requires the collection and publication of continuous, differentiated data at all academic career levels, in particular concerning the participation of women and men. The same applies to other dimensions of diversity — as far as legally permissible," the guidelines say.

Asked for more details, the DFG told *Nature* that it doesn't have any specific rules on how to collect data on cultural or ethnic representation — besides recommending anonymous surveys that pay careful attention to data protection. The funder says that it is up to universities to decide how to conduct such surveys and what data to collect, and that it doesn't have an overview of these activities. "Universities and research institutions can and should already take measures to promote diversity in view of the evidence that is already available," a spokesperson said, referring to surveys by the DZHW and others. But the DFG also said that it is not aware of any universities already implementing such measures.

Last September, the German Rectors' Conference launched a federally funded initiative called Diversity at German Universities. Details are limited, but the aim is to fund some universities to examine what hurdles and barriers need to be overcome to enable more diversity and inclusivity, with results due to be presented in 2024; monitoring diversity is also a goal.



How UK science is failing Black researchers — in nine stark charts

For Joshua Kwesi Aikins, a political scientist and human-rights activist at the University of Kassel, who worked on the Afrozensus for CFE and is joining zedela, the time to collect information on race and on multiple forms of discrimination has already come. However, he adds, such data must be collected only with the full understanding and participation of the groups who are being asked for the information, so that the exercise is done in good faith.

"History has led Germany to be rightfully cautious about racial classification. But German institutions need to act," he says. "A strikingly male-dominated academe perpetuates racist, sexist and classist dynamics that add to structural hurdles on the way to what remains for too many a precarious academic career."

Aikins and Wagner are part of a three-year research network funded by the DFG and running until 2025, which aims to define new categories for data

collection. This could include self-reported ethnicity, among other facets of identity. The group, coordinated by Anne-Kathrin Will, who studies migration at the Humboldt University of Berlin, and Linda Supik, a sociologist at Leibniz University in Hanover, is working with non-governmental and civil-society organizations.

Whether these efforts will change how German employers or universities collect race or ethnicity data remains to be seen. The DFG says that, in future, researchers wanting funding will need to address how they are promoting diversity, and that it plans to develop individual measures for this "over a period of approximately five years". It also notes that it is developing a survey to understand "needs and expectations" related to its promotion of diversity, but says it cannot comment further.

At the Ruhr University Bochum, meanwhile, Karle says work is being done to improve the understanding of concerns over discrimination. Last July, for instance, Hans Alves, a psychologist at the university who specializes in social cognition, ran a survey of more than 4,000 students asking about their gender, migration status, religion and experiences of various kinds of discrimination. It did not ask about race or ethnicity. The results have not yet been published, but Alves told *Nature* the survey showed that 10% of students with a parent born outside Germany said they had been the target of discrimination because they were perceived as 'foreign' or 'non-white'. Karle says the university is now thinking about conducting a similar project across all of its academic staff.

Nature **616**, 22-24 (2023)

doi: https://doi.org/10.1038/d41586-023-00955-9

Additional reporting by Quirin Schiermeier.

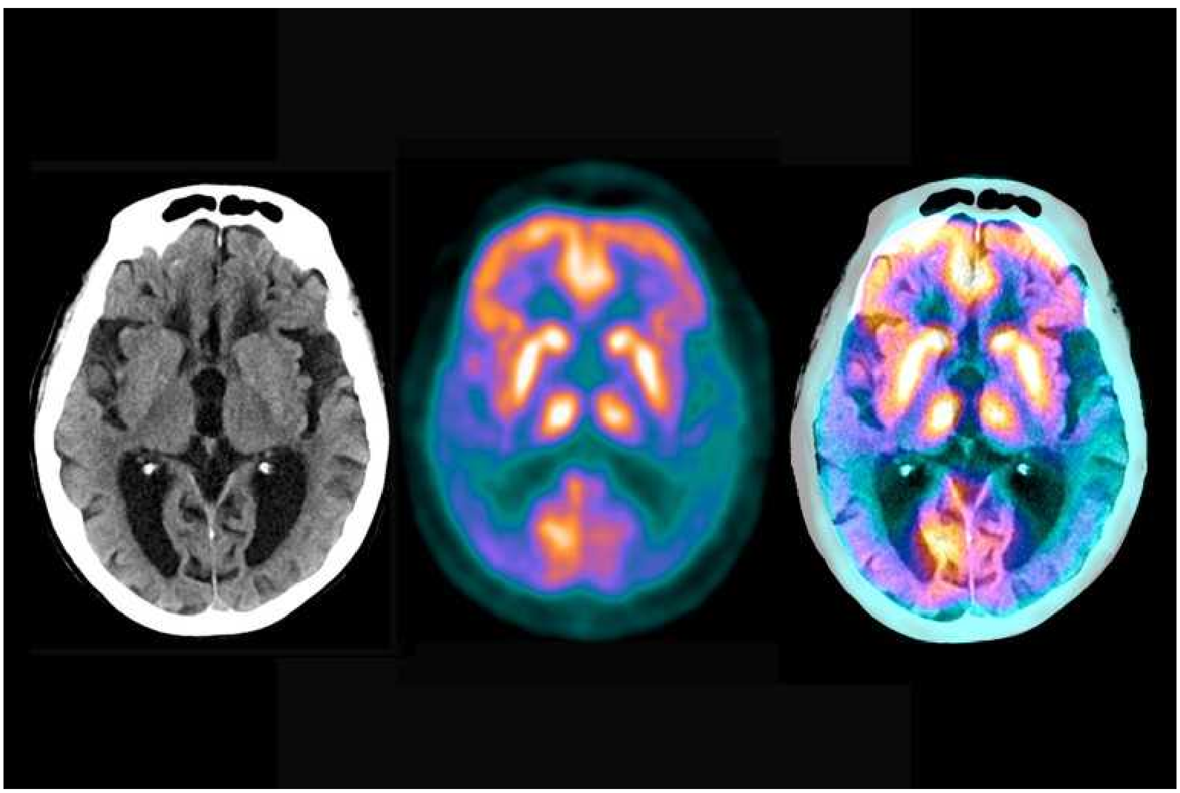
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00955-9

- NEWS FEATURE
- 04 April 2023

Conquering Alzheimer’s: a look at the therapies of the future

Researchers are looking to drug combinations, vaccines and gene therapy as they forge the next generation of treatments for the condition.

- Alison Abbott



A brain scan reveals the extent of damage caused by Alzheimer’s disease. Credit: Zephyr/Science Photo Library

When neurologist Reisa Sperling stepped up to receive her lifetime achievement award at an international Alzheimer's conference last December, she was more excited about the future than about celebrating the past.

What thrilled Sperling, who won the award for her work on clinical trials of Alzheimer's treatments, was a sense of hope, which has been conspicuously missing from research into the disease for many years. Most other attendees felt the same.

Just a few months before the meeting, researchers had announced that an antibody drug called lecanemab clearly lowered the amount of amyloid protein plaques — a tell-tale sign of the disease — in the brains of participants in a clinical trial, and slowed their cognitive decline.

Sperling, who runs a laboratory at Harvard Medical School in Boston, Massachusetts, was buoyant as she gripped the microphone tightly. After spending more than 30 frustrating years in Alzheimer's research, she said, there was finally proof that she and her colleagues were on the right track. "But still, it isn't enough," she said.

In the trial, treatment led to a 25% slowing of decline, enough to give participants a few extra months of independent living[1]. "But actually conquering a destructive disease that affects tens of millions of people worldwide is a different story," she says.

What's more, lecanemab, marketed in the United States as Leqembi, makes for a tough treatment regime. It has to be infused through a vein by a nursing professional. And because the drug can cause potentially life-threatening brain swelling and bleeds, people taking it have to be monitored regularly.



Could drugs prevent Alzheimer's? These trials aim to find out

A similar antibody, aducanumab, was approved by the US Food and Drug Administration in 2021, but the decision generated controversy because the drug's clinical trial had not shown unambiguous benefit.

Despite these wrinkles, the lecanemab results consolidated the hope that Alzheimer's might eventually be preventable, if treatment is given early enough. The success also raised another possibility: this and future drugs could be used in combination to tackle different stages of the disease, which are often governed by different molecules. Few expect a single therapy to be the answer. But combination-therapy trials are expensive and complicated, because each drug has to be tested both on its own and with its partner. Pharmaceutical companies can be wary of hitching their product to another, in case the combination fails and casts a shadow over their drug.

The mood of greater confidence among those in the field is not only down to the success of anti-amyloid antibodies. The cupboard is stacked with potential new therapies — and abandoned drugs that are now being dusted off.

Promising combinations

Alzheimer's disease has a long, silent beginning. First, plaques — clumps of sticky amyloid-β proteins — start to gather in the brain. They quickly become surrounded by immune cells called microglia that try, but ultimately fail, to nibble them away. The plaques grow in size and number but go

unnoticed for years, even decades, until they cause another protein — called tau — to accumulate to toxic levels and spread around the brain in tangles. Scientists are still working out exactly how this chain of events occurs, but cognitive symptoms begin to emerge only when it is well under way. The severity of the symptoms correlates with the extent of the tau tangles.

Individual therapies directed at tau have so far not performed well in trials[2]. But scientists think that tau-busting drugs might work better when paired with anti-amyloid therapy.

“We know that amyloid somehow drives the accumulation of pathological tau, which then spreads through the brain like wildfire,” says Randall Bateman, a neurologist at Washington University in St. Louis, Missouri. “So we think it makes sense to remove the substance that is fuelling the flames, while we try to put out the fire of tau tangles.”



Randall Bateman (left) and colleagues will run a trial of anti-tau and anti-amyloid antibodies.Credit: Matt Miller/Washington University School of Medicine

Bateman and his colleagues began planning such a trial in 2015, but it became feasible only recently, after the first hints emerged that amyloid therapies might prove effective. Last year, they launched an international trial known as Tau NexGen. They are recruiting 168 participants, all of whom are likely to develop Alzheimer's disease at a young age — typically in their 30s, 40s or 50s — because they have a mutation in a gene that leads them to overproduce amyloid-β.

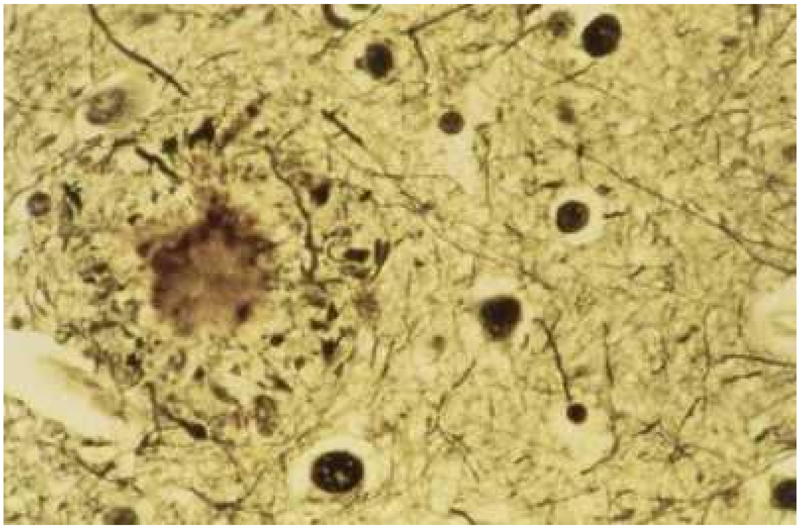
Participants are being divided into two groups, according to whether they already have symptoms of dementia or are expected to develop symptoms in the next ten years (these people usually become symptomatic at around the same age as the parent from whom they inherited the mutation).

All participants will be given lecanemab and a tau-reducing antibody, but in different orders. Those without symptoms will receive the anti-tau antibody E2814 for a year, then have lecanemab added; the symptomatic group will receive lecanemab for six months and then have E2814 added. The researchers running the trial hope they can use this set-up to learn about optimal treatment combinations.

More anti-tau drugs will eventually be included in the study, and the first results of the trial are expected after 2027.

Tau NexGen is the first, and so far only, ongoing clinical trial of a combination therapy for the condition. A similar, US-based, trial[3] is being planned for sporadic, late-onset Alzheimer's disease, which affects older people and accounts for the overwhelming majority of cases. The US National Institutes of Health (NIH) is expected to decide in the next few months whether it will co-fund this effort, called the ATP trial, as a public–private partnership with drug companies. If it does, recruitment could start next year.

Many pharmaceutical and biotechnology companies are developing anti-tau therapies, some as antibodies, others using other small molecules or newer genetic approaches to block the production of pathological forms of tau. ATP trial co-leader Adam Boxer, a neurologist at the University of California, San Francisco, says that several of these companies have formally expressed interest in participating.



[Alzheimer’s drug slows mental decline in trial — but is it a breakthrough?]

Like Tau NexGen, this will be a type of prevention trial. Participants will have few or no detectable symptoms, but will have evidence from blood tests and scans that their brains contain plaques and early signs of tau tangles. There will be about 900 participants in 6 arms, who will receive one of two tau therapies either alone or in combination with lecanemab, lecanemab alone, or a placebo.

The research teams hope that anti-tau treatment will boost lecanemab’s modest benefits — and that, in a virtuous circle, by reducing plaque burden, lecanemab will create better conditions for anti-tau therapy to work.

Key to the trials is a range of sensitive new biomarkers — measurements of the brain or blood that give a read-out of disease status. Brain scans monitor the presence and severity of amyloid plaques and tau tangles; tests on blood or cerebrospinal fluid measure many other molecules in the chain of pathology, such as different forms of amyloid and tau. Researchers expect that the extensive molecular and clinical data that they produce will help to uncover more about the mechanisms of Alzheimer’s. “Evidence so far points to tau as the initiator of Alzheimer’s symptoms, disabilities and eventually death,” says Boxer. “But this hypothesis needs to be tested in humans.”

Trials of combination therapies have some drawbacks: they are complex and costly to manage. Boxer estimates that the ATP trial will run to many

hundreds of millions of dollars, despite the new biomarkers, which make trials a lot more efficient.

Antibodies themselves make for expensive therapies. Lecanemab will be marketed at US$26,500 for a year’s treatment. Aducanumab (sold as Aduhelm in the United States) was originally priced at $56,000 for a year’s treatment, but the makers halved the price after public outcry.

The drugs are also inconvenient for patients because they have to be infused every few weeks. Data from clinical trials suggest that life-long therapy will be required to keep Alzheimer’s at bay, says neurologist Paul Aisen, at the University of Southern California in San Diego, who is a leader of the US Alzheimer’s Clinical Trials Consortium. “The disease appears to rebound when infusions stop,” he says. Because long-term antibody treatment is impractical, he says, “we think it might make sense to maintain a low amyloid state with an oral drug that blocks generation of the peptide” once the antibodies have removed the plaques.

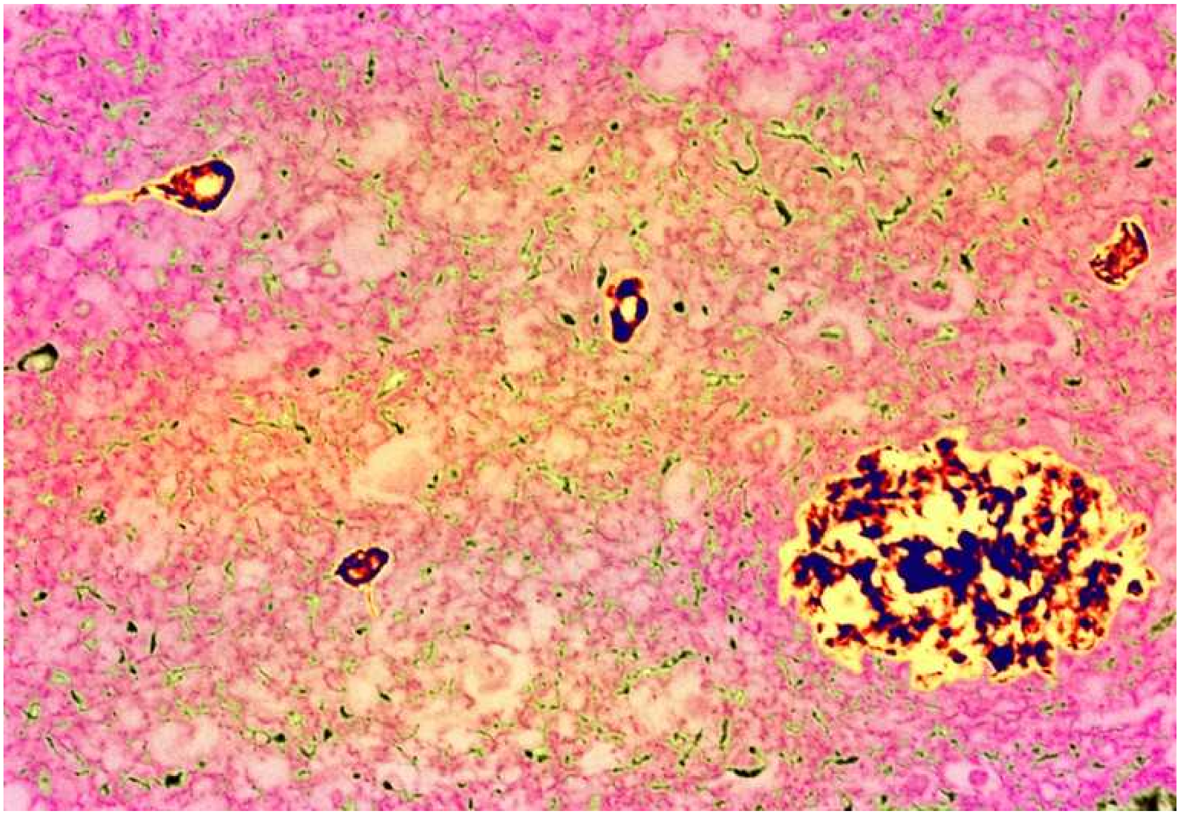
Such compounds do exist. Beginning around 2010, researchers trialled a suite of oral drugs that aimed to reduce amyloid in the brain by tuning down the activities of one of two enzymes, β-secretase and γ-secretase, which are key to its production. But clinical trials of these drugs all flopped[4], and interest in them dried up — until they got a second chance.

Other ingredients

In 2018, a group of drug companies agreed to do something unusual in their normally secretive industry. They decided to share confidential clinical data from six failed trials with each other and a select group of specialists who had been brought together by the Alzheimer’s Association, a lobby group for patients’ rights based in Chicago, Illinois.

The association wanted to learn as much as possible from the disastrous clinical trials, each of which tested a different inhibitor of the β-secretase enzyme. None of the drugs had shown benefit — and worse, many had toxic side effects, including, in some cases, worsened cognition. Rather than let the trial data gather dust behind closed doors, the association wanted them

discussed in detail. The goal, says the association’s chief scientific officer, Maria Carrillo, was “to help the field understand more about the disease-related biologies targeted by these investigational drugs”.



Classical features of Alzheimer’s: a plaque (large clump) and tau tangles (small clumps).Credit: Simon Fraser/Science Photo Library

The panel’s review paper, published in 2021, suggests that the disease might have been too advanced in trial participants for this class of drug to improve symptoms, and that lower doses could avoid side effects[5].

Aisen thinks it should be possible to use these drugs at low doses to stop plaques coming back — once existing plaques have been cleared with antibodies.

A batch of clinical trials targeting γ-secretase also failed. But rather than abandon the target, researchers have been working on a more nuanced approach. Instead of completely blocking the enzyme — a blunt tool that

contributes to the toxic side effects observed in trials — they hope to change its behaviour.

One such modulator, a compound developed in an academic collaboration[6], will be tested in an early clinical trial this year. The drug, which can be taken orally, causes the enzyme to chop amyloid up into shorter proteins, which are non-toxic and could even be protective. The trial will be sponsored by a new start-up company, Acta Pharmaceuticals in Boston, Massachusetts, and funded by the NIH.

Most drugs being considered for combination trials target amyloid or tau. But there are earlier-stage approaches that seek to improve the brain's natural immune-defence mechanisms against Alzheimer's. Once again, researchers have learnt a lot from families with genetic mutations that predispose them to Alzheimer's.

The mutation in question is in a gene called *TREM2*, which makes a molecule that sits on the surface of the brain's immune-system warriors, microglia. "Tuning up microglia could make them more efficient in clearing plaques or prevent amyloid pathology from spreading," says neuroscientist Christian Haass at the Ludwig Maximilian University of Munich in Germany, "particularly if the plaque load is first reduced with anti-amyloid therapy." He is planning experiments in mice to test how an antibody that binds TREM2 and activates microglia might work if it is taken alongside anti-amyloid therapy. A similar antibody is in an early-phase clinical trial as a solo treatment.

Vaccines, genes and plasma

More approaches to stemming Alzheimer's are making their way into clinical trials. Researchers aim to deliver useful molecules to the brain — through vaccines, viral vectors or blood transfusions.

Like the enzyme drugs, vaccines are being reinvented, long after the first clinical trial of an anti-amyloid vaccine was halted in 2002, when brain inflammation was observed in some participants.



[FDA approves Alzheimer’s drug lecanemab amid safety concerns]

There are now several anti-tau and anti-amyloid vaccines in preparation or in early-phase clinical trials. They comprise snippets of the tau or amyloid-β proteins, chosen and packaged to avoid serious inflammatory responses. They are designed to stimulate the brain’s immune system to recognize and destroy the full versions of the proteins, and were intended mainly to prevent disease or to slow the progression of early disease. Scientists are even trying to develop vaccines to attack both tau and amyloid-β.

Other researchers are betting on gene therapy to conquer forms of Alzheimer’s caused by gene mutations.

Distinct forms of the *APOE* gene, which encodes a protein involved in fat metabolism, affect the risk of Alzheimer’s disease in different ways. The *APOE4* allele is linked to increased risk, whereas the *APOE2* allele decreases it. For its open-label trial, in which participants all received the therapy, Lexeo Therapeutics, based in New York City, recruited 15 volunteers with mild Alzheimer’s symptoms who have 2 copies of the *APOE4* gene. They wanted to test whether delivering the *APOE2* gene variant would mitigate the harmful effect of the higher-risk version. So they attached the gene to a viral vector, which they then injected directly into the volunteers’ spinal fluid.

The gene graft seems to have been successful. Lexeo reported last year that the *APOE2* gene has been detected in the spinal fluid of some participants up to a year after injection, that no serious side effects have been observed so far and that the participants' tau levels have gone down. It's too early to assess whether the disease progression has slowed as a result, but participants will be monitored until 2028.

Gene therapy won't be suitable for everyone, because known gene mutations determine only a small proportion of cases of Alzheimer's. But the concept of replacement therapy has been taken up by others. The company Alkahest, based in San Carlos, California, has completed a small clinical trial testing whether factors in the blood of young people could replace those lost in the ageing process[7].

Some researchers have found success with lower-tech methods. For example, a large, carefully controlled trial showed that 18 months of either aerobic exercise or stretching stalled cognitive decline in people with mild cognitive impairment, underlying the value of maintaining an exercise regime during pharmacological treatment.

It is too early to say which, if any, of these potential new therapies will pan out. Most researchers think that treatment will need to be personalized: individuals at different disease stages will need different therapies. "It's great to see so many plausible approaches being pursued," says Aisen. "But we have a long way to go."

Still, the range of options going through clinical testing is encouraging, says Sperling. "And our new glimmer of success is moving us forward."

Nature **616**, 26-28 (2023)

doi: https://doi.org/10.1038/d41586-023-00954-w

References

1. Van Dyck, C. H. *et al. N. Engl. J. Med.* **388**, 9–21 (2023).

2. Imbimbo, B. P., Balducci, C., Ippati, S. & Watling, M. *Neural Regen. Res.* **18**, 117–118 (2023).

3. Aisen, P. S. *et al. J. Prev. Alzheimers Dis.* **8**, 306–312 (2021).

4. Bazzari, F. H. & Bazzari, A. H. *Molecules* **27**, 8823 (2022).

5. McDade, E. *et al. Nature Rev. Neurol.* **17**, 703–714 (2021).

6. Rynearson, K. D. *et al. J. Exp. Med.* **218**, e20202560 (2021).

7. Hannestad, J. *et al. J. Alzheimers Dis.* **81**, 1649–1662 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00954-w

Opinion

- **Marine heatwaves need clear definitions so coastal communities can adapt** [03 Apr 2023]
  • Clearly communicating baselines for assessing ocean warming is essential for understanding extreme events and how they will affect marine ecosystems and livelihoods in the future.

- COMMENT
- 03 April 2023
- Correction 05 April 2023

Marine heatwaves need clear definitions so coastal communities can adapt

Clearly communicating baselines for assessing ocean warming is essential for understanding extreme events and how they will affect marine ecosystems and livelihoods in the future.

- Dillon J. Amaya ,
- Michael G. Jacox,
- Melanie R. Fewings,
- Vincent S. Saba,
- Malte F. Stuecker,
- Ryan R. Rykaczewski,
- Andrew C. Ross,
- Charles A. Stock,
- Antonietta Capotondi,
- Colleen M. Petrik,
- Steven J. Bograd,
- Michael A. Alexander,
- Wei Cheng,
- Albert J. Hermann,
- Kelly A. Kearney &
- …
- Brian S. Powell

Show authors



Coral bleaching is developing in Japan’s largest coral reef, Sekiseishoko, which lies between Ishigaki and Iriomote islands. Credit: BJ Warnick/Kyodo/Newscom/Alamy

Marine heatwaves devastate ecosystems and the coastal communities that rely on them. Weeks, months or years of unusually warm waters can bleach corals, spur harmful algal blooms and wipe out seaweeds. They might kill or strand marine animals and disrupt food webs and fisheries[1]. Billions of US dollars are lost to such events around the world each year[2].

For example, in 2013, an area of water in the northeast Pacific Ocean more than three times the size of Texas, known as The Blob, warmed by nearly 3°C. Over 18 months, these warm waters spread across the entire west coast of North America, from the Gulf of Alaska to the tip of the Baja Peninsula in Mexico. Seabirds starved and stocks of Pacific cod collapsed. Tuna moved north, as far as Alaska. Humpback whales drawn towards the coast became entangled in fishing nets. Mysterious creatures, such as glowing tropical sea pickles, or pyrosomes, arrived in northern waters.

Ocean scientists are striving to better understand such phenomena, and whether climate change is making marine heatwaves more frequent and

more intense. But right now the field has a problem: the definitions and communications describing what a marine heatwave is are confusing.



[Five priorities for a sustainable ocean economy]

This is because scientists and the media use the term ‘marine heatwave’ to refer to two different things: extreme conditions compared with historical temperatures, and extreme conditions compared with an evolving ‘new normal’ of rising temperatures owing to climate change. The difference matters. Each definition leads to different estimates of the properties and trends of future marine heatwaves. Coastal communities need to understand what is unfolding if they are to adapt.

Because the precise meaning remains unclear, miscommunication and misunderstanding are rife. Duelling definitions mean that headlines such as ‘marine heatwaves are getting more frequent’ and ‘marine heatwaves are not getting more frequent’ can be simultaneously true, for reasons that are not obvious to the public. This breakdown in communication is having real-world consequences for decision makers tasked with managing oceans in the coming decades.

Here we — the members of the US National Oceanic and Atmospheric Administration (NOAA) Marine Ecosystem Task Force — propose how to

define basic terms to clarify what kinds of ocean warming the world has in store. (See Supplementary information for a list of co-signatories.)

Baselines matter

At its simplest, a marine heatwave is a period of extreme ocean temperatures. These anomalies can grow and fade rapidly over a few days or weeks, or slowly over months or years. Temperatures are considered extreme if they exceed a threshold (often the 90th percentile, or 90% of expected fluctuations) above ‘normal’ temperatures for that region and season. The question is, what is ‘normal’?

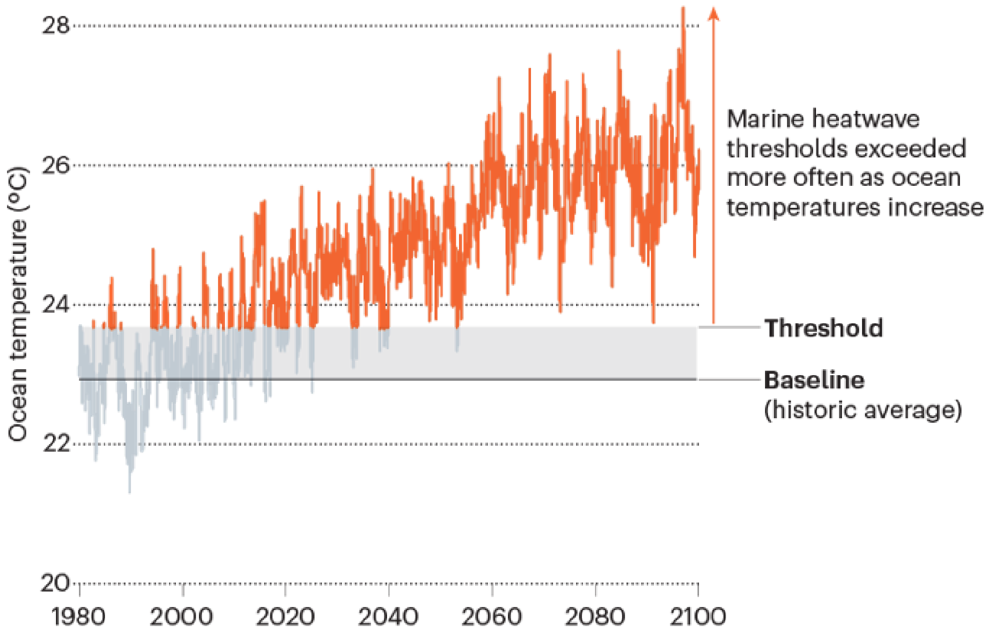
In one definition of marine heatwaves, ‘normal’ refers to typical variations in water temperature for a given area and time of year, based on a set period in the past (often spanning 30 years or more). This makes sense in many ways. But, this fixed baseline approach doesn’t take into account that oceans are getting warmer over time because of climate change. Ocean surface waters have warmed at an average rate of 0.06 °C per decade since pre-industrial times, compared with 0.1 °C per decade for surface air temperature over land; both rates are rising[3]. Using the fixed baseline definition, over the coming decades, we would see ever-more-frequent and more intense marine heatwaves until the oceans warm so much that they reach a ‘perpetual heatwave’ state (see ‘Marine heatwaves: duelling definitions’).

MARINE HEATWAVES: DUELLING DEFINITIONS

Assessing spikes of extreme ocean temperatures using different baselines* paints two different pictures for the future as the climate warms. Coastal communities need to know which definition is being used so they can plan.

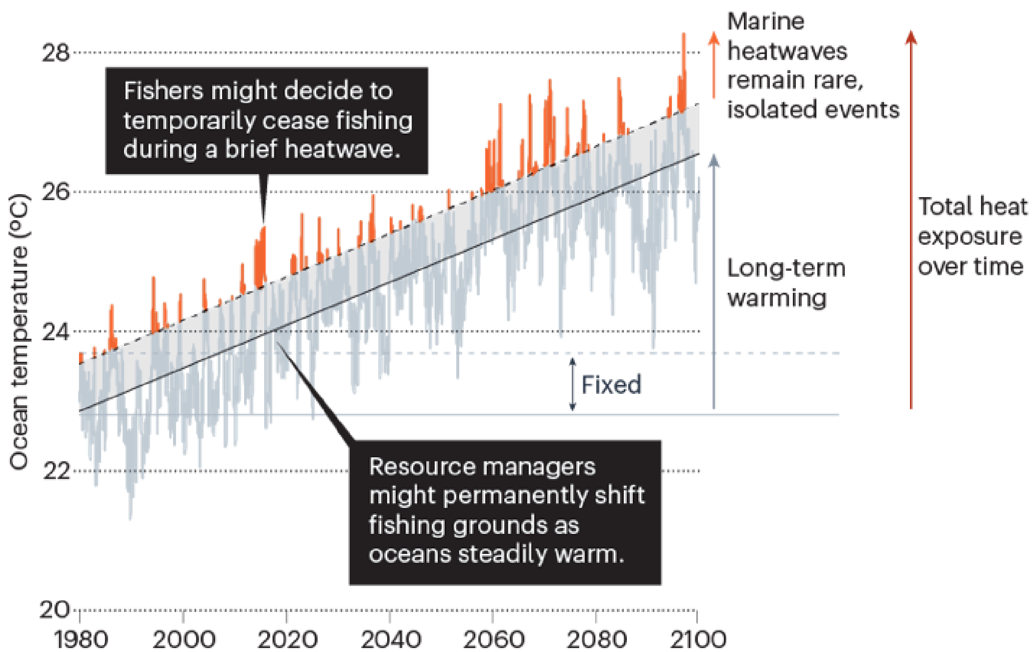
Fixed baseline

Measuring heat relative to historical temperatures makes sense for tracking coral bleaching, for example, but says little about patterns of future extremes.



Shifting baseline

Defining marine heatwaves relative to increasing average temperatures helps resource managers to distinguish temporary changes and long-term trends.



*Baselines and thresholds are illustrative only; seasonal variations are not considered for simplicity.

©nature

Source: Source: J. E. Kay *et al. Bull. Am. Meteorol. Soc*. **96**, 1333–1349 (2015)

The second definition uses a shifting baseline of 'normal' conditions for any given patch of ocean. In other words, it accounts for long-term ocean warming separately and effectively subtracts that from the total. From this perspective, in the future, ocean temperature anomalies would have to be much warmer than those derived from a fixed historic baseline to qualify as a marine heatwave. That might seem unfair. But this definition retains the idea that a heatwave should be an exceptional event in time and space — a central tenet of the original, qualitative definition of marine heatwaves proposed in 2016[4].



Sustainable small-scale fisheries can help people and the planet

Each definition results in different interpretations of frequency, intensity, duration and spatial extent of marine heatwaves. For example, one prominent 2019 study using a fixed baseline approach predicted that large swaths of the world's oceans would be warmer than the fixed historical threshold for heatwaves by 2100[5]. The shifting baseline approach, however, predicts more-subtle changes in marine heatwave intensity and frequency by 2100 and only for certain regions, such as the northeast Pacific[6]. The exact values depend on the models used.

Clear messaging about the nature of ocean temperature change over short and long timescales is crucial, so that time and resources are appropriately allocated. Resource managers, industries and coastal communities have different options for dealing with slow, steady warming versus rapid, temporary temperature change. For example, the fishing industry might apply emergency strategies, such as temporarily ceasing fishing, to cope with short-term disruptions. Adapting to long-term warming might instead require shifting to other fishing grounds or target species and obtaining new permits and gear[7].

However, the messaging is currently far from clear. For example, the US 2022 State of the Ecosystem (SOE)[8] report for New England coastal waters, which used a fixed baseline definition, reported that the Gulf of Maine experienced marine heatwaves for more than 80% of the year in 2021. In recent decades these waters have experienced some of the strongest warming trends on the planet, at more than 0.3 °C per decade[9].



In 2019, a harmful algal bloom developed on Lake Hopatcong in New Jersey.Credit: Rick Loomis/New York Times/Redux/eyevine

Without that context, however, a marine-resource manager equipped with only the 2022 SOE report could justifiably assume that rapid-fire high-heat

events separated by short spurts of ‘normal’ now dominate the Gulf of Maine, and begin to adjust their decision-making accordingly. In reality, however, this fixed-baseline statistic does not distinguish whether this region is experiencing more frequent episodic temperature extremes on top of the long-term warming.

The US team that leads the annual SOE report issued by the NOAA Northeast Fisheries Science Center has recognized this issue and, starting with the 2023 SOE report, it will detail marine heatwaves using the shifting baseline approach. Temperature changes relative to a fixed baseline will also be reported, but those excursions will not be described as marine heatwaves.

Two definitions can be useful

Both the fixed and shifting baseline methods have merits, and some scientists suggest that the choice of which one to use should be tailored to the application of interest[10]. For example, the fixed baseline approach might suit some studies of marine-organism physiology[5]. Many species, such as corals, are severely affected when water temperature reaches a specific threshold.



UN high seas treaty is a landmark – but science needs to fill the gap

Similarly, on land, researchers often judge atmospheric heatwaves relative to fixed historical baselines. Human-health impacts due to extreme heat are also linked to temperature thresholds, and infrastructure (such as houses, air-conditioning units and highways) is designed to tolerate certain temperature ranges. The chance of hitting a ‘perpetual heatwave state’ against a fixed baseline is much smaller on land than at sea, because the temperature anomalies are larger[11]. On land, where surfaces gain and lose heat more rapidly than the ocean, heatwaves tend to be about 5 °C warmer than usual temperatures, much larger than the trend due to climate change, compared with about 1–2 °C for marine heatwaves.

But there are also problems with the fixed baseline approach when it comes to marine heatwaves. Suggesting that these events will become more extreme and frequent in the future implies that ecological impacts will also become more intense, frequent and widespread. This is unlikely to always be true.

The relatively slow march of global warming and the more rapid shock of an intense marine heatwave will elicit different physiological and behavioural responses from many marine species[12]. For example, those with short life cycles, such as market squid (*Doryteuthis opalescens*) or sardine species, might have some capacity to adapt to a slow temperature increase but not necessarily to rapid heat shock. By contrast, some corals might recover from acute temperature impacts but not chronic heat exposure.



A humpback whale caught in a fishing net in Mallorca, Spain.Credit: Cordon Press/Alamy

The ways in which marine life moves around the ocean will also differ in response to different types of warming[13]. Some species of fish caught for commercial fishing can shift location rapidly in response to temporary bouts of warming. For example, Pacific bluefin tuna (*Thunnus orientalis*) off the coast of California moved 1,000 kilometres north in response to The Blob in 2013–15. It is unclear, however, whether such sudden changes in behaviour will become more common in other locations in the future. Instead, in slowly warming waters, displacement of many species might be steady and persistent rather than abrupt and temporary.

Public perceptions of risks from marine heatwaves might also be damaged if people continue to hear that these events will become more extreme in the future, using a fixed baseline. If everything is extremely warm all of the time, then the term 'extreme' loses its meaning. The public might become

desensitized to the real threat of marine heatwaves, potentially leading to inaction or a lack of preparedness.

There are other compelling physical and ecological arguments in favour of the shifting baseline approach. It is consistent with how people think of a temporary surge in ocean temperature. And it makes it easier to see the connections to other forms of climate variability (such as the El Niño–Southern Oscillation) when long-term warming is removed from the equation.

Clarify what's coming

Both long-term warming and short-term, discrete extreme events are important to quantify, but they should not be called the same thing. We suggest the following unambiguous language that allows scientists to clearly differentiate and communicate the various events and impacts. And we encourage the community to discuss and agree on the use of such or similar terms.



Save reefs to rescue all ecosystems

First, use the phrase 'long-term temperature trends' to describe the relatively slow changes in ocean temperature that occur over decades or longer.

Second, only use the term ‘marine heatwave’ with the shifting baseline definition, to describe ocean temperature changes that are transient and extremely warm relative to the expected conditions for a given place and time, as defined by an evolving, recent climatological reference period.

Third, use the term ‘total heat exposure’ to describe the combination of long-term warming and marine heatwaves. This terminology is applicable to studies of impacts on ecosystems that are based on a fixed temperature threshold.

An analogy can be drawn between rising ocean temperatures and sea levels. Over decades, the average height of the global ocean creeps up slowly, centimetre-by-centimetre, forcing coastal infrastructure to gradually adapt to keep buildings and people dry. Marine heatwaves are like storm surges: relatively short-lived and intense changes in ocean height that lead to acute impacts. Both storm surges and long-term sea-level rise contribute to the total ocean height, just as the combination of long-term temperature trends and marine heatwaves drive total heat exposure.

By adopting our suggested ocean temperature communication framework, scientists will be able to better equip marine resource decision makers to assess and prepare for risks associated with different types of ocean temperature change.

Nature **616**, 29-32 (2023)

doi: https://doi.org/10.1038/d41586-023-00924-2

Updates & Corrections

- **Correction 05 April 2023**: An earlier version of this article gave an incorrect affiliation for Ryan R. Rykaczewski.

References

1. Smale, D. A. *et al. Nature Clim. Chang.* **9**, 306–312 (2019).

2. Smith, K. E. *et al. Science* **374**, eabj3593 (2021).

3. IPCC. *Summary for Policymakers*. In: *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. *et al.*) 3−32 (Cambridge Univ. Press, 2021).

4. Hobday, A. J. *et al. Prog. Oceanogr.* **141**, 227–238 (2016).

5. Oliver, E. C. J. *et al. Front. Mar. Sci.* **6**, 734 (2019).

6. Xu, T. *et al. Nature Commun.* **13**, 7396 (2022).

7. Fisher, M. C., Moore, S. K., Jardine, S. L., Watson, J. R. & Samhouri, J. F. *Proc. Natl Acad. Sci. USA* **118**, e2014379117 (2021).

8. Northeast Fisheries Science Center. *State of the Ecosystem 2022: New England* (NOAA, 2022).

9. Pershing, A. J. *et al. Science* **350**, 809–812 (2015).

10. Oliver, E. C. J. *et al. Annu. Rev. Mar. Sci.* **13**, 313–342 (2021).

11. Frölicher, T. L. & Laufkötter, C. *Nature Commun.* **9**, 650 (2018).

12. Provost, M. M. & Botsford, L. W. *Oikos* **2022**, e08909 (2022).

13. Pinsky, M. L. *et al. Science* **360**, 1189–1191 (2018).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00924-2

Work

- **Split-site doctorates are helping to build Africa's research base** [03 Apr 2023]
  • University partnerships that let PhD students split their time between two countries and keep jobs at home have many benefits.
- **Cultivating cannabis in the shadow of a volcano** [03 Apr 2023]
  • Agriculturist Rohan McDonald is rebuilding a medical-cannabis laboratory after an eruption destroyed his team's work.

- CAREER FEATURE
- 03 April 2023

Split-site doctorates are helping to build Africa’s research base

University partnerships that let PhD students split their time between two countries and keep jobs at home have many benefits.

- Rachel Nuwer

Find a new job



Peter Waiswa argues that joint PhD programmes benefit not just African research but also global science.Credit: Grace Amoit

Peter Waiswa was working as an assistant district health officer and a physician for Uganda’s Ministry of Health when he heard about an intriguing opportunity: a chance to earn a PhD in public health through a joint programme sponsored by Makerere University in Kampala and the Karolinska Institute in Stockholm. The programme, funded primarily by the Swedish government’s International Development Cooperation Agency, aimed to support Uganda’s development by investing in a new generation of scientists. “Most PhDs focus on the individual, but in this programme, the focus was on building institutional capacity at Makerere,” Waiswa says. “The research you do was also meant to be aligned to national development priorities.”

The opportunity to gain exposure to a new culture and education system appealed to Waiswa, who already had a master’s degree in public health. In 2006, he applied and was accepted. He spent three months in each year of the four-year programme in Sweden, and worked jointly with one Ugandan and two Swedish supervisors to hone his doctoral research on health systems and policy in Africa. His intention, Waiswa says, was always to remain in Uganda and contribute to his country’s science and public-health programmes.

While enrolled, he got a job with Makerere University, where he founded and now coordinates the Centre of Excellence for Maternal Newborn & Child Health, which aims to promote evidence-based policymaking and health-care interventions for mothers and newborns in Uganda and neighbouring countries. Still a visiting researcher at the Karolinska Institute, he publishes at least a dozen papers each year, supervises several master’s and PhD students at Makerere and has brought in more than 30 grants totalling more than US$20 million. Waiswa’s affiliation with the Karolinska, the experience he gained in Sweden and his international professional network have all helped him to achieve personal success, he says, and have benefited Makerere and Uganda more generally, too.

“Do these programmes advance Africa’s agenda? I think they do,” Waiswa says.



Decolonizing science toolkit

Most African universities have plenty of staff with master's-level qualifications but few with a PhD, says Pauline Essah, director of research and programmes at Education Sub Saharan Africa (ESSA), a non-profit organization registered in the United Kingdom and Ghana that promotes high-quality education in sub-Saharan Africa. The continent needs another one million doctoral scholars to catch up with the international average for researchers per capita (see go.nature.com/3ixvjyt), and Africa's dearth of PhD holders is already creating a bottleneck in training new scholars to fill this need. "Universities are where you should be generating knowledge and coming up with solutions to major problems, but we are massively short" of the expertise needed, Essah says. "Economies and countries in Africa suffer as a result."

That's where split-site PhD programmes such as the one Waiswa participated in can be crucial. By "tapping into expertise, mentorship and supervision from outside the continent", such programmes can help to "get our PhD numbers up to the level where they're eventually self-sufficient," Essah says.

These programmes usually aim to build overall research capacity in Africa because, by design, they encourage students to stay in their home countries and get an academic, civil society or government job that allows them to

conduct locally relevant research. “We don’t remove people from their home setting by putting them in the US or UK for a few years, by the end of which they’re estranged from their home environment,” says Stefan Swartling Peterson, a physician and global-health researcher at the Karolinska Institute, who was Waiswa’s PhD supervisor.

The institute, Waiswa points out, also benefits from the partnership. Its faculty members are able to bring in extra funding through joint proposals with Makerere colleagues, and Karolinska academics and students can easily collaborate with Makerere scholars and obtain access to field sites in Uganda. “Their professors and students get exposure to real-world issues,” Waiswa says. The partnership has also improved the Karolinska’s “credibility and visibility and impact, especially in global health”.

Despite the pluses for both sides, there are not enough of these programmes to meet demand in Africa, and those that do exist tend to be tethered to the funding whims of Western governments or institutions. New split-site PhD programmes that Sweden was developing with Rwanda and Tanzania, for example, are now “on the chopping block”, Peterson says, owing to national budget cuts. But failure to support Africa’s ability to conduct research and innovate on its own is short-sighted, he says, and will cost the world more in the long run.

No single model

Split-site PhDs are sometimes referred to as joint, double or sandwich degrees, and every programme is different. Some are aimed at recent master’s graduates, whereas others are specifically designed for university academics or government employees who do not have a PhD. In some split-site programmes, those who already have a professional position can keep their job, increasing the chances that they will remain in their country after graduating.

Although these programmes are usually supported by Western governments, organizations or universities, they are often established by researchers from partner institutions who are already working together, or academics with knowledge of both countries involved. Bernt Lindtjørn, a public-health

researcher at the University of Bergen in Norway, chose Hawassa University in Ethiopia as the focus for a joint programme that he helped to establish at Bergen (one now funded by the Norwegian government) because he was born in Ethiopia and had worked there as a physician. Likewise, Peterson helped to launch the collaboration between Makerere University and the Karolinska Institute after conducting research at Makerere himself.

Some split-site PhDs result in a joint degree conferred by both partner universities; others award a PhD from one institution. There is no standardized formula for what the programmes themselves entail. Some require equal time to be spent at both institutions, whereas others involve a shorter stint abroad to take a few courses, access resources — such as laboratories, libraries and computational facilities — and analyse results. Some don't require African students to travel at all, and instead send academics from the partner institution to Africa or make courses available online.

Women, especially, can benefit from programmes that do not involve a heavy travel component, Essah says, because long periods away from their families might discourage them from applying in the first place. Some of the split-site programmes that Essah was involved in setting up and running from 2009 to 2018, involving the University of Cambridge, UK, and several African universities, included an optional mid-programme break for women with young children. "That's another plane ticket you have to purchase, but it ensures women who have caregiving duties can go and see how their children and families are doing," she says.



The idea of a PhD with just a short stint in Bergen, Norway, appealed to Hiwot Abera Areru.Credit: Hiwot Abera

Hiwot Abera Areru says she chose to enrol in the split-site PhD programme in public health at Hawassa University — where she is now an assistant professor — and the University of Bergen, specifically because of the limited time she would be required to spend in Norway. “I was in the middle of establishing my own family,” she says. “I preferred this joint PhD because of the prestigiousness of the University of Bergen over local universities, but also because it would have been difficult for me to go totally abroad.”

Areru spent just two months in Norway, but thanks to the relationships she made there, she’s now undertaking a joint postdoc, also split between Hawassa and Bergen. “Joint programmes should be implemented in all universities in Africa,” Areru says. “We need more of them, especially for females who need to be near their families but also want to fulfil their career ambitions.”

Obstacles to success

Split PhD programmes are not without shortcomings. For starters, few of them exist. Often, says Essah, organizations “would rather fund three master’s-level scholarships than fund a PhD”.

It can be difficult, too, to find those split-site PhD programmes that do exist. No centralized database compiles or compares them, so potential students often depend on word of mouth or their own initiative. Essah and her colleagues at ESSA are trying to fill this gap with a scholarship hub that lists vetted opportunities for African students, including split-site PhDs. The news organization Al-Fanar Media compiles a similar list of opportunities open to students in North African countries.



Developing an inclusive culture at South Africa’s research institutions

Students who wish to enrol in one of these programmes usually need to attend an African university that specifically offers such a partnership, although some exceptions do exist. For example, the UK Foreign, Commonwealth and Development Office sponsors split-site scholarships for PhD students from low- or middle-income countries that are members of the Commonwealth of Nations — including 20 in Africa — to conduct research for 12 months at a UK university.

Limitations exist, too, in the type of PhDs offered on these programmes. Usually, they focus on addressing social needs, meaning that opportunities are more plentiful in fields such as public health, agriculture and engineering than in, say, astronomy, particle physics or wildlife biology. Language can also limit the scope of classes that PhD students can take, Waiswa says. Academics at the Karolinska Institute, for example, taught public-health classes in English but most other subjects in Swedish.

The logistics can be daunting as well. It's usually up to students to find their own supervisors at both locations, and they can also face difficulties in trying to meet the requirements of two cultures and institutions. "Sometimes, you wind up doing more work than other students," Waiswa says.

Some split-site PhDs require faculty members at African universities to continue teaching while they pursue their doctorate, and these programmes might not provide any extra funding, on top of their salary, to cover living expenses. Often, in such cases, "PhDs are not completed in time because the candidates can't focus just on their research," says Beate Schindler-Kovats, director of the Nairobi office of the German Academic Exchange Service. Schindler-Kovats sees just three or four applications per year for split-site PhDs in Germany, she says, because "the incentives are not good enough" to compete with just doing a PhD abroad.



Promoting palaeontology across Sudan

Academics who want to create split-site PhD programmes frequently face challenges, too. Trying to get everyone on the same page about academic credits and scholarly standards is “a total nightmare”, says Dan Haydon, director of the Glasgow Centre for International Development at the University of Glasgow, UK. Haydon and his colleagues in the School of Biodiversity, One Health and Veterinary Medicine tried for years to set up a split-site master’s programme with an African university but were ultimately unsuccessful. Unless an external funder steps in, he says, Western universities are usually reluctant to invest time and effort in creating a programme that won’t be a major source of income in the short term.

Quality-control concerns can also be an issue. The University of Bergen, for example, pushed back at first on Lindtjørn’s idea of conferring a degree jointly with an Ethiopian university, because administrators at Bergen could not directly control standards at the partner institution. Lindtjørn reassured them that graduates would have to satisfy all the requirements of any Norwegian institution. “It’s the same, identical game, and thereby we kept the quality,” he says.

Others have found more-creative workarounds. Tom Kawula, director of the Paul G. Allen School for Global Health at Washington State University in Pullman, pushed for six years to set up a joint PhD programme in biomedical sciences with the University of Nairobi, but struggled to get around the “roadblock” of having to make a one-to-one comparison between courses in the United States and Kenya, he says. The solution finally arrived when he and his colleagues realized that rather than requiring a “specific checklist” of perfectly transferable courses, they could instead focus on scholarly outputs, such as the requirement that students publish two first-author papers in high-quality journals. “That was the biggest break in the dam,” he says, and the joint programme is now set to launch in August.

Reciprocal benefits

Split-site PhDs are often touted as a way to prevent the brain drain in Africa, but the goal shouldn’t be to make sure every student returns home, Essah says. “Sometimes diaspora Africans are given a hard time and are told ‘You

should be back in Africa,'" she says. "But the reality is, we cannot all be in Africa, and it's a personal choice."

Sometimes, other factors make the choice for scientists. Eskindir Loha Shumbullo, for example, is an Ethiopian scholar who got his PhD in epidemiology at the University of Bergen in 2013. He returned to his job at Hawassa University, where he helped to launch the split PhD programme with Bergen. But when ethnic-based violence broke out in Ethiopia in 2018, Eskindir was forced to look for a job abroad. He ended up back at the University of Bergen as a postdoc. "This postdoc position is kind of downgrading my status," he says. "But I came to Norway for the sake of peace."



[Measuring societal impact: how to go beyond standard publication metrics]

No single organization seems to track what happens to split-site PhD scholars throughout their careers, which is something Essah says needs to change. "A lot of this data and information is just sitting on funders' computers," she says. However, anecdotal numbers suggest that these programmes do seem to be effective at building and retaining talent in Africa. Waiswa says that about 90% of the 34 scholars who graduated from his programme remained in Uganda, and almost all of the 100-plus split-site PhDs and postdocs that Essah supported through the University of Cambridge's programmes are back in Africa. Most are conducting research,

Essah says, and some have become heads of departments, faculties or universities.

More Western universities and governments should recognize, too, that the benefits of these programmes are about more than the individual researchers themselves, Haydon says. Such partnerships foster opportunities for collaborative research that will advance knowledge and solve problems in and beyond Africa. “This isn’t a charitable thing, this is about getting the research done right,” Haydon says. “The rewards are absolutely massive, and it’s really sad that it’s so hard for university managers to see that.”

There are signs that Africa is starting to take the lead in establishing and supporting these partnerships. In 2015, a coalition of African governments created the Regional Scholarship and Innovation Fund, which supports split-site PhD programmes aimed at training high-quality doctoral students and raising the calibre of African research. As the first pan-African science fund of its kind, it exemplifies the continent’s increasing commitment to fostering growth and development through science and innovation. “African governments are seeing the need for PhD-level expertise,” Essah says.

The fund, which receives support from the World Bank and some countries outside Africa, for now focuses on five priority areas: food security; energy; mining, mineral and materials engineering; information and computer technology; and climate change. Since 2018, it has supported more than 250 African scholars in science, technology, engineering and mathematics, and thousands more have applied — a testament, Essah says, to the interest in these programmes and “the reality” that the world needs more of them.

Nature **616**, 207-209 (2023)

doi: https://doi.org/10.1038/d41586-023-00956-8

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00956-8

- WHERE I WORK
- 03 April 2023

Cultivating cannabis in the shadow of a volcano

Agriculturist Rohan McDonald is rebuilding a medical-cannabis laboratory after an eruption destroyed his team's work.

- James Mitchell Crow



Rohan McDonald is a tropical agriculture researcher at the Orange Hill Agricultural Biotechnology Centre, St Vincent and the Grenadines. Credit: Micah B Rubin for *Nature*

The Orange Hill Agricultural Biotechnology Centre that I lead in St Vincent and the Grenadines in the southern Caribbean is only six kilometres or so from the La Soufrière volcano. So when the volcano explosively erupted on 9 April 2021, we had to get out quick.

We knew back in early 2021 that La Soufrière could erupt any time, so we had moved our research material to a safe place. But the eruption destroyed all our greenhouses and shade houses and badly damaged our concrete buildings.

Where I'm standing had been one of our largest shade houses, and the tray in front of me once contained cannabis plantlets. Our facility aims to contribute to the government's plan to launch a medical-cannabis export industry.

I started out at Orange Hill in the plant tissue-culture laboratory in 2000 and still have a passion for it. Tissue culture involves growing plantlets from cuttings under sterile, controlled conditions. Our country sometimes struggles with crop diseases and needs the clean plant material that tissue culture provides.

Cannabis was once grown illegally up in the mountains. It is now cultivated legally, but there is a very high percentage of male plants, which don't produce the cannabinoid-rich flower needed for medicinal products.

Using tissue culture, we can start with a female plant and guarantee that all progeny are female and disease-free. Despite the delays caused by the eruption, we now have a tissue-culture protocol that works for all the cannabis varieties we have collected.

The government mantra is to 'build back better', so we will have two large temperature-controlled greenhouses equipped with solar panels, and a secure lab for our cannabis research, designed to be more resistant to eruption damage. The reconstruction should be complete by August.

Nature **616**, 212 (2023)

doi: https://doi.org/10.1038/d41586-023-00957-7

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00957-7

Research

- **The medieval Moon unveils volcanic secrets** [05 Apr 2023]
  • Innovative use of medieval musings about the Moon has revealed that volcanic eruptions coincided with abrupt, global-scale cooling events. The approach is exciting from the perspective of climate scientists and historians alike.
- **Beating the break-even point with a discrete-variable-encoded logical qubit** [22 Mar 2023]
  • The lifetime of a qubit in a circuit quantum electrodynamics architecture can exceed the break-even point through quantum error correction, representing a key step towards scalable quantum computing.
- **Lunar eclipses illuminate timing and climate impact of medieval volcanism** [05 Apr 2023]
  • Analysis of contemporary reports of total lunar eclipses, combined with aerosol model simulations and tree-ring-based climate proxies, allowed greater precision in dating of the occurrence of stratospheric volcanic eruptions during the High Medieval Period.
- **Four ways blue foods can help achieve food system ambitions across nations** [22 Feb 2023]
  • A study proposes four ways in which foods sourced in aquatic environments can contribute to healthier, more environmentally sustainable and equitable food systems, and examines the relevance of these ambitions to nations.
- **Spatial epigenome–transcriptome co-profiling of mammalian tissues** [15 Mar 2023]
  • The authors present two technologies for spatially resolved, genome-wide, joint profiling of the epigenome and transcriptome by cosequencing chromatin accessibility and gene expression, or histone modifications and gene expression on the same tissue section at near-single-cell resolution.
- **Spatial multiomics map of trophoblast development in early pregnancy** [29 Mar 2023]
  • A multiomics single-cell atlas of the human maternal–fetal interface including the myometrium, combining spatial transcriptomics data with chromatin accessibility, provides a comprehensive analysis of cell states as placental cells infiltrate the uterus during early pregnancy.

- **RHOJ controls EMT-associated resistance to chemotherapy** [22 Mar 2023]
  • RHOJ regulates epithelial-to-mesenchymal-transition-associated resistance to chemotherapy by enhancing the response to replicative stress and activating the DNA damage response, enabling tumour cells to rapidly repair DNA lesions induced by chemotherapy.
- **From primordial clocks to circadian oscillators** [22 Mar 2023]
  • X-ray, cryo-EM and kinetic studies of the circadian oscillator KaiBC from the photosynthetic bacterium Rhodobacter sphaeroides shed light on the evolution of self-regulating oscillators.
- **Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1** [22 Mar 2023]
  • Using cryo-electron microscopy, the molecular architecture and catalytic mechanism of action of the fungal β-1,3-glucan synthase FKS1 are determined.

- NEWS & VIEWS FORUM
- 05 April 2023

The medieval Moon unveils volcanic secrets

Innovative use of medieval musings about the Moon has revealed that volcanic eruptions coincided with abrupt, global-scale cooling events. The approach is exciting from the perspective of climate scientists and historians alike.

- Andrea Seim ,
- Eduardo Zorita &
- Anne Lawrence-Mathers

THE PAPER IN BRIEF

• Volcanic eruptions inject large volumes of sulfur dioxide into the atmosphere that are converted into aerosols in the stratosphere[1].

• These aerosols produce volcanic dust that can reduce incoming solar radiation, altering Earth's surface temperatures, precipitation and atmospheric circulation[2].

• Identifying past volcanic eruptions can therefore help to clarify the timing and nature of climate events, but current methods produce conflicting results.

• In a paper in *Nature*, Guillet *et al.*[3] report an approach that interprets medieval accounts of lunar eclipses to date volcanic eruptions that could have shaped a key climatic transition.

ANDREA SEIM & EDUARDO ZORITA: A fresh take on an old challenge

Climate scientists usually identify past volcanic eruptions by measuring the acidity and amount of volcanic ash in cores drilled from polar ice, or by inferring abrupt temperature changes in tree-ring records. However, these sources sometimes disagree, because the location, intensity and timing of eruptions can produce varying results, as can circulation of the atmosphere. Guillet and colleagues' approach offers an independent — and perhaps even more direct — source of information about the timing of volcanic eruptions, which could resolve some of these disagreements.

During a total lunar eclipse, the Moon is fully in Earth's shadow. A dark Moon indicates that volcanic aerosols are highly abundant in the stratosphere, whereas a reddish Moon suggests that they are scarce (Fig. 1). Guillet *et al.* examined historical accounts of lunar eclipses from the High Medieval Period (1100–1300), and estimated the abundance of volcanic aerosols from the descriptions of the colour and luminosity of the Moon. They used this information to refine the timing of a cluster of volcanic eruptions that occurred during this period, and which had previously been identified using ice-core measurements[4]. The authors found that seven of these eruptions generated aerosols that could have had a role in the transition from the Medieval Climate Anomaly (around 850–1250) to the Little Ice Age (around 1300–1850).

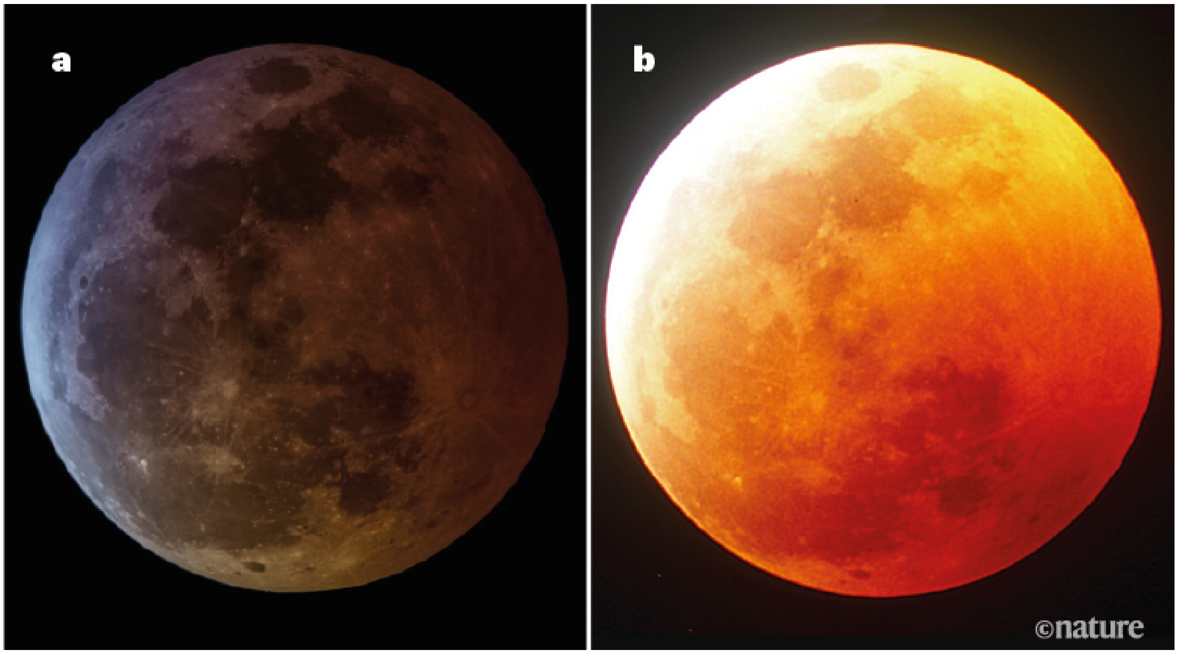


Figure 1 | The colours of total lunar eclipses. a, A dark Moon during a lunar eclipse indicates that Earth's stratosphere contains a large amount of volcanic aerosols. This photograph was taken in January 2019, roughly one month after the eruption of Anak Krakatau in Indonesia. **b**, A bright reddish Moon suggests that volcanic aerosols are scarce, as shown in this photograph from January 2001. Guillet *et al*.[3] used historical accounts of the colours of lunar eclipses to infer the climatic impact of volcanic eruptions in the High Medieval Period (1100–1300).Credit: Giuseppe Donatiello (CC0 1.0)

The strength of Guillet and co-workers' study lies in the precision with which the authors estimated the timing of volcanic eruptions — pinpointing the year, and even in some cases the month, of the event. The authors compared their findings with modern global aerosol measurements, climate model simulations and satellite observations to link five dark and two reddish lunar eclipses to major eruptions during the High Medieval Period. They found that the dark lunar eclipses were observable for 3–20 months after an eruption.



Read the paper: Lunar eclipses illuminate timing and climate impact of medieval volcanism

They then examined tree-ring records sensitive to summer temperatures in the Northern Hemisphere, in which an unusually cold summer is indicated by reduced wood formation. By combining these records with climate simulations, the authors further refined the timing of five eruptions and showed that they had a pronounced impact on the climate. The remaining eruptions seem to have had less effect.

Guillet and colleagues’ findings will improve our understanding of how much volcanic eruptions contributed to the onset of the Little Ice Age compared with other factors, including a period of low solar activity between 1280 and 1340. Climate models suggest that a volcano-induced cooling in surface temperatures can lower ocean temperatures and expand sea ice[5]. This can, in turn, stretch the effects of a volcano over several decades, and could even plunge the climate system into a different state — a prospect that has relevance for modern warming as a result of anthropogenic greenhouse-gas emissions. Although the climatic changes caused by volcanic and anthropogenic emissions differ, understanding how the climate system reacts to different types of perturbation will also help to refine the next generation of climate models[6]. And getting the timing right is crucial, which is why innovative approaches such as that of Guillet and colleagues are so valuable.

ANNE LAWRENCE-MATHERS: Making sense of historical records

Guillet and colleagues' findings are underpinned by hundreds of records chronicling 187 lunar eclipses between 1100 and 1300. Their approach to linking these events to the climatic impact of volcanic eruptions hinges on descriptions of the colour of the Moon. Such information is scarce in these records, except in Christian sources, in which the colour and brightness of the Moon is documented in 36 of the eclipses studied. But the authors faced difficulties in dating these eclipses because the sources are not always consistent — a challenge related to medieval church bureaucracy.



The aberrant global synchrony of present-day warming

In the medieval Christian world, lunar phases and eclipses were used for setting the dates of religious festivals and timing medical treatments — or for determining whether a darkened Moon was a sign from God. New astronomical tools and theories were therefore valued highly. Calculating the lunar calendar accurately was important to Christians because the Moon touched on so many of their activities.

Such studies certainly became more accurate when Latin translations of Islamic astronomical texts and planetary tables were made available in

Europe from around 1150[7]. However, these texts used an entirely different calendar and referenced dates that were meaningful only for their original compilers. They also described astronomical observations that were made far from northern Europe. Early translations were literal and offered little help to users unfamiliar with the Islamic calendar.

The resulting struggles are clear in the writings of a monk known as John, who chronicled sunspots and a solar eclipse from his post at Worcester Cathedral in England at the start of the High Medieval Period. John made detailed records of these astronomical phenomena, but still failed to understand the Islamic methods for calculating lunar eclipses[8,9]. Nevertheless, the planetary tables were valued, and — once more-sophisticated translations and recalculations had been produced — they formed the basis of textbooks on astronomy. One such book was written by Johannes de Sacrobosco in around 1220, and immediately became widely used (Fig. 2). How, then, could there be continuing doubts about the timing and identification of lunar eclipses?

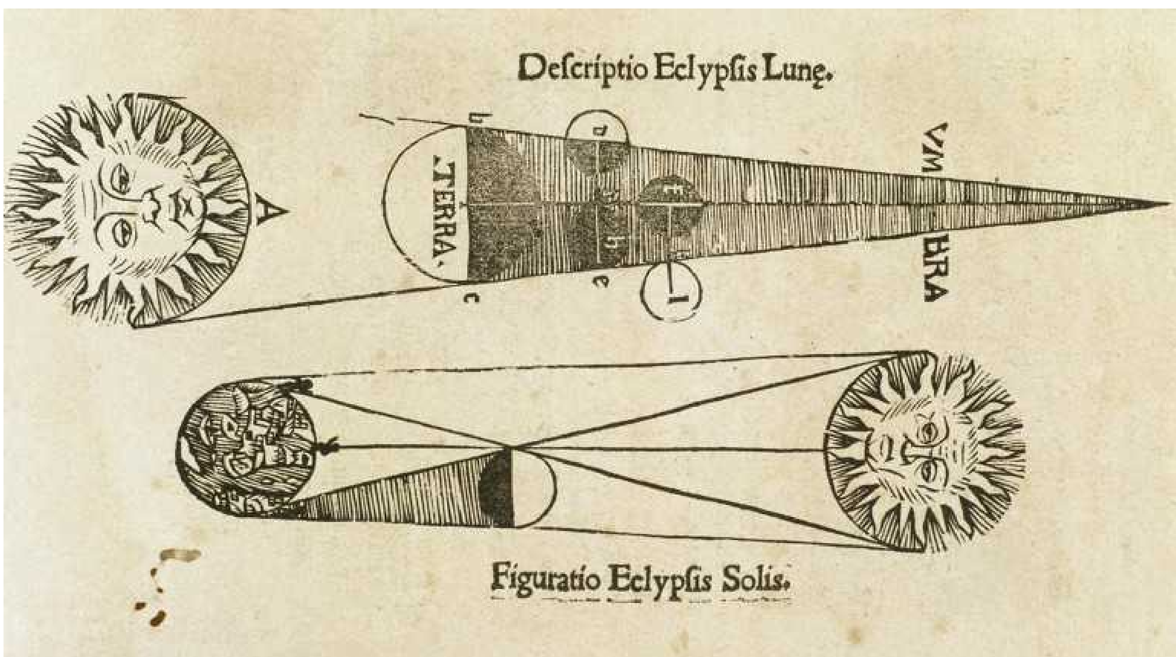


Figure 2 | A diagram of solar and lunar eclipses, from a printed edition of Johannes de Sacrobosco's astronomical text (around 1220). Credit: Prima Archivo/Alamy

Central to the answer is the fact that the solar year did not align exactly with a year in the Julian calendar. Sacrobosco argued that the accumulated error had reached some 11 days by around 1230 and that Easter was therefore being celebrated on the wrong day[10]. Church leaders were aware of the problem, but feared there would be serious resistance to changing the calendar[11]. The timing of phases of the Moon was therefore out of step with the calendars of major religious institutions.

Church chroniclers faced a dilemma. Some saw a darkening of the Moon as a clear sign of an eclipse, whereas others interpreted it differently. The problem was exacerbated by the fact that a given eclipse might be more clearly visible in some parts of Europe than others, and that information took time to be transmitted between regions. For these reasons, the lack of agreement between medieval chroniclers on the timing of total eclipses of the Moon should not be seen as the result of religious censorship or ignorance — they were simply days out of sync. In this context, Guillet and co-workers' accomplishment in overcoming this problem to link historical sources to climate events is all the more impressive.

Nature **616**, 38-40 (2023)

doi: https://doi.org/10.1038/d41586-023-00850-3

References

1. Rampino, M. R. & Self, S. *Nature* **310**, 677–679 (1984).
2. Robock, A. *Rev. Geophys.* **38**, 191–219 (2000).
3. Guillet, S. *et al. Nature* **616**, 90–95 (2023).
4. Sigl, M. *et al. Nature* **523**, 543–549 (2015).
5. Miller, G. H. *et al. Geophys. Res. Lett.* **39**, L02708 (2012).
6. Merlis, T. M., Held, I. M., Stenchikov, G. L., Zeng, F. & Horowitz, L. W. *J. Clim.* **27**, 7781–7795 (2014).

7. Al-Khalili, J. *Pathfinders: The Golden Age of Arabic Science* 67–78 (Lane, 2010).

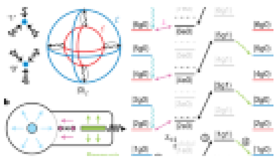
8. Nothaft, C. P. E. *Early Sci. Med.* **20**, 187–208 (2015).

9. Lawrence-Mathers, A. E. *J. Mediev. Hist.* **39**, 255–274 (2013).

10. Moreton, J. *Viator* **25**, 229–244 (1994).

11. Nothaft, C. P. E. *Scandalous Error: Calendar Reform and Calendrical Astronomy in Medieval Europe* (Oxford Univ. Press, 2018).

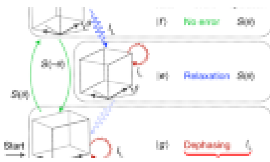
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-00850-3



Protecting a bosonic qubit with autonomous quantum error correction

10 February 2021

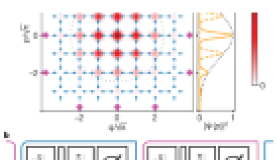
Jeffrey M. Gertler, Brian Baker, … Chen Wang



Error-corrected gates on an encoded qubit

08 June 2020

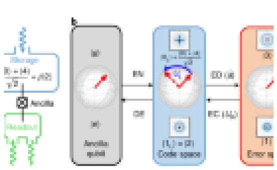
Philip Reinhold, Serge Rosenblum, … Robert J. Schoelkopf



Quantum error correction of a qubit encoded in grid states of an oscillator

19 August 2020

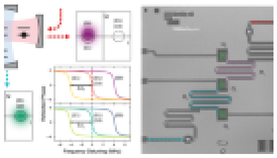
P. Campagne-Ibarcq, A. Eickbusch, … M. H. Devoret



[Quantum error correction and universal gate set operation on a binomial bosonic logical qubit]

11 February 2019

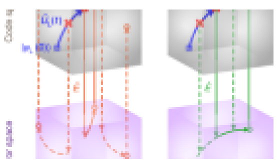
L. Hu, Y. Ma, … L. Sun



[Experimental demonstration of continuous quantum error correction]

28 April 2022

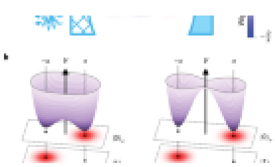
William P. Livingston, Machiel S. Blok, … Irfan Siddiqi



[Error-transparent operations on a logical qubit protected by quantum error correction]

11 May 2020

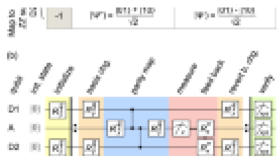
Y. Ma, Y. Xu, … L. Sun



[Exponential suppression of bit-flips in a qubit encoded in an oscillator]

16 March 2020

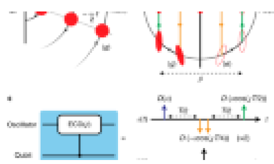
Raphaël Lescanne, Marius Villiers, … Zaki Leghtas



[Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits]

15 August 2019

Christian Kraglund Andersen, Ants Remm, … Christopher Eichler



[Fast universal control of an oscillator with weak dispersive coupling to a qubit]

17 October 2022

Alec Eickbusch, Volodymyr Sivak, … Michel H. Devoret

- Article
- Open Access
- Published: 22 March 2023

Beating the break-even point with a discrete-variable-encoded logical qubit

- Zhongchu Ni,
- Sai Li,
- Xiaowei Deng,
- Yanyan Cai,
- Libo Zhang,
- Weiting Wang,
- Zhen-Biao Yang,
- Haifeng Yu,
- Fei Yan,
- Song Liu,
- Chang-Ling Zou,
- Luyan Sun ,
- Shi-Biao Zheng ,
- Yuan Xu &
- …
- Dapeng Yu

Show authors

Nature **volume 616**, pages 56–60 (2023)

- 8895 Accesses
- 1 Citations
- 74 Altmetric
- Metrics details

Abstract

Quantum error correction (QEC) aims to protect logical qubits from noises by using the redundancy of a large Hilbert space, which allows errors to be detected and corrected in real time[1]. In most QEC codes[2,3,4,5,6,7,8], a logical qubit is encoded in some discrete variables, for example photon numbers, so that the encoded quantum information can be unambiguously extracted after processing. Over the past decade, repetitive QEC has been demonstrated

with various discrete-variable-encoded scenarios[9,10,11,12,13,14,15,16,17]. However, extending the lifetimes of thus-encoded logical qubits beyond the best available physical qubit still remains elusive, which represents a break-even point for judging the practical usefulness of QEC. Here we demonstrate a QEC procedure in a circuit quantum electrodynamics architecture[18], where the logical qubit is binomially encoded in photon-number states of a microwave cavity[8], dispersively coupled to an auxiliary superconducting qubit. By applying a pulse featuring a tailored frequency comb to the auxiliary qubit, we can repetitively extract the error syndrome with high fidelity and perform error correction with feedback control accordingly, thereby exceeding the break-even point by about 16% lifetime enhancement. Our work illustrates the potential of hardware-efficient discrete-variable encodings for fault-tolerant quantum computation[19].

Main

One of the main obstacles for building a quantum computer is environmentally induced decoherence, which destroys the quantum information stored in the qubits. The errors caused by decoherence can be corrected by repetitive application of a quantum error correction (QEC) procedure, whereby the logical qubit is encoded in a high-dimensional Hilbert space, such that different errors project the system into different orthogonal subspaces and thus can be unambiguously identified and corrected without disturbing the stored quantum information. In conventional QEC schemes[1,9], the code words of a logical qubit are formed by two highly symmetric entangled states of several physical qubits encoded with some discrete variables. The past two decades have witnessed remarkable advances in experimental demonstrations of this kind of QEC code in different systems, including nuclear spins[5,6], nitrogen-vacancy centres in diamond[10,20], trapped ions[7,11,21,22,23], photonic qubits[24], silicon spin qubits[25] and superconducting circuits[12,13,14,15,16,26,27]. However, in these experiments, the lifetime of the logical qubit still needs to be greatly extended to reach that of the best available physical component, which is regarded as the break-even point for judging whether or not a QEC code can benefit quantum information storage and processing.

An alternative QEC encoding scheme is to use the large space of an oscillator, which can be used to encode either a continuous-variable or a discrete-variable qubit[28,29,30,31,32]. Both types of code can tolerate errors due to loss and gain of energy quanta, enabling QEC to be performed in a hardware-efficient way. Circuit quantum electrodynamics (QED) systems[18] represent an ideal platform for realizing such encoding schemes: the break-even point has been exceeded in two breakthrough experiments[33,34] by distributing the quantum information over an infinite-dimensional Hilbert space of a continuous-variable-encoded photonic qubit, but the code words of this photonic qubit are not strictly orthogonal. This inherent restriction can be overcome with discrete-variable encoding schemes, whereby the code words of a logical qubit are encoded with mutually orthogonal Fock states of an oscillator. This feature, together with their intrinsic compatibility with error-correctable gates[35,36] and their usefulness in logically connecting modules in a quantum network[37], makes such discrete-variable qubits promising in fault-tolerant quantum computation. These advantages can be turned into practical benefits in real quantum information processing only when the lifetime of the encoded logical qubits is extended beyond the break-even point, which, however, remains an elusive result, although enduring efforts have been made towards this goal[17,32].

Here, we demonstrate the exceeding of the QEC break-even point by real-time feedback correction for a discrete-variable photonic qubit in a microwave cavity, whose code words remain mutually orthogonal and can be unambiguously discriminated. The dominant error, single photon loss, of the logical qubit is mapped to the state of a Josephson junction-based nonlinear oscillator that is dispersively coupled to the cavity and serves as an auxiliary qubit, realized with a continuous pulse involving an ingeniously tailored comb of frequency components. As the driving frequencies aim at the error space where a photon loss event occurs, perturbations on the logical qubit are highly suppressed when it remains in the encoded logical space. Another intrinsic advantage of this error syndrome detection is that the continuous driving protects the system from the auxiliary qubit's dephasing noise. We demonstrate this procedure with the lowest-order binomial code and extend the stored quantum information lifetime 16% longer than the best physical qubit, encoded in the two lowest Fock states and referred to as the Fock qubit. A more important characteristic associated

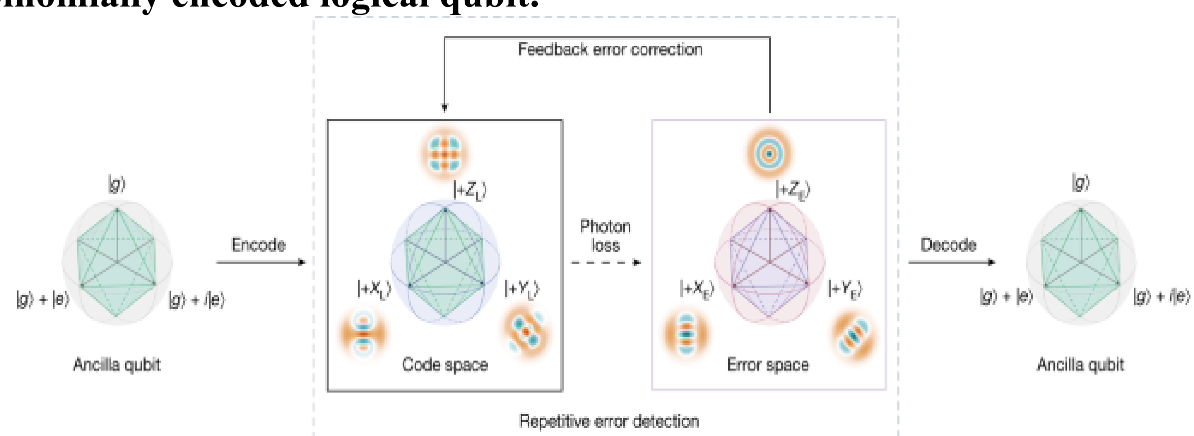
with this error-detecting procedure is that neither the logical nor the error space needs to have a definite parity, which allows the implementation of QEC codes that can tolerate losses of more than one photon.

The key stages of a QEC procedure are encoding the quantum information to the logical qubit from the auxiliary qubit, the error syndrome measurement, the real-time error correction of the system depending on the measurement output and the decoding process to read out the quantum information stored in the logical qubit. Our logical qubit is realized in a three-dimensional microwave cavity, and the dominant decoherence to combat is the excitation loss error. The logical qubit is encoded with a binomial code[8], with the code words:

$$\begin{array}{r}| {0}_{{\rm{L}}}\rangle =(| 0\rangle +| 4\rangle)/\sqrt{2},\\ | {1}_{{\rm{L}}}\rangle =| 2\rangle ,\end{array}$$

where the number in each ket denotes the photon number in the cavity. The binomial code is a typical stabilizer QEC code: when the single-photon-loss error occurs, the quantum information is projected into the error space spanned by $\{\left|{0}_{{\rm{E}}}\right\rangle =\left|3\right\rangle ,\left|{1}_{{\rm{E}}}\right\rangle =\left|1\right\rangle \}$, with the photon number parity acting as the error syndrome to distinguish these two spaces. A general QEC protection of quantum information stored in the bosonic system is illustrated in Fig. 1. After correctly measuring the photon number parity and applying the corresponding correction operations in real time, quantum information stored in the cavity can be recovered.

Fig. 1: Schematic of the QEC procedure with the lowest-order binomially encoded logical qubit.



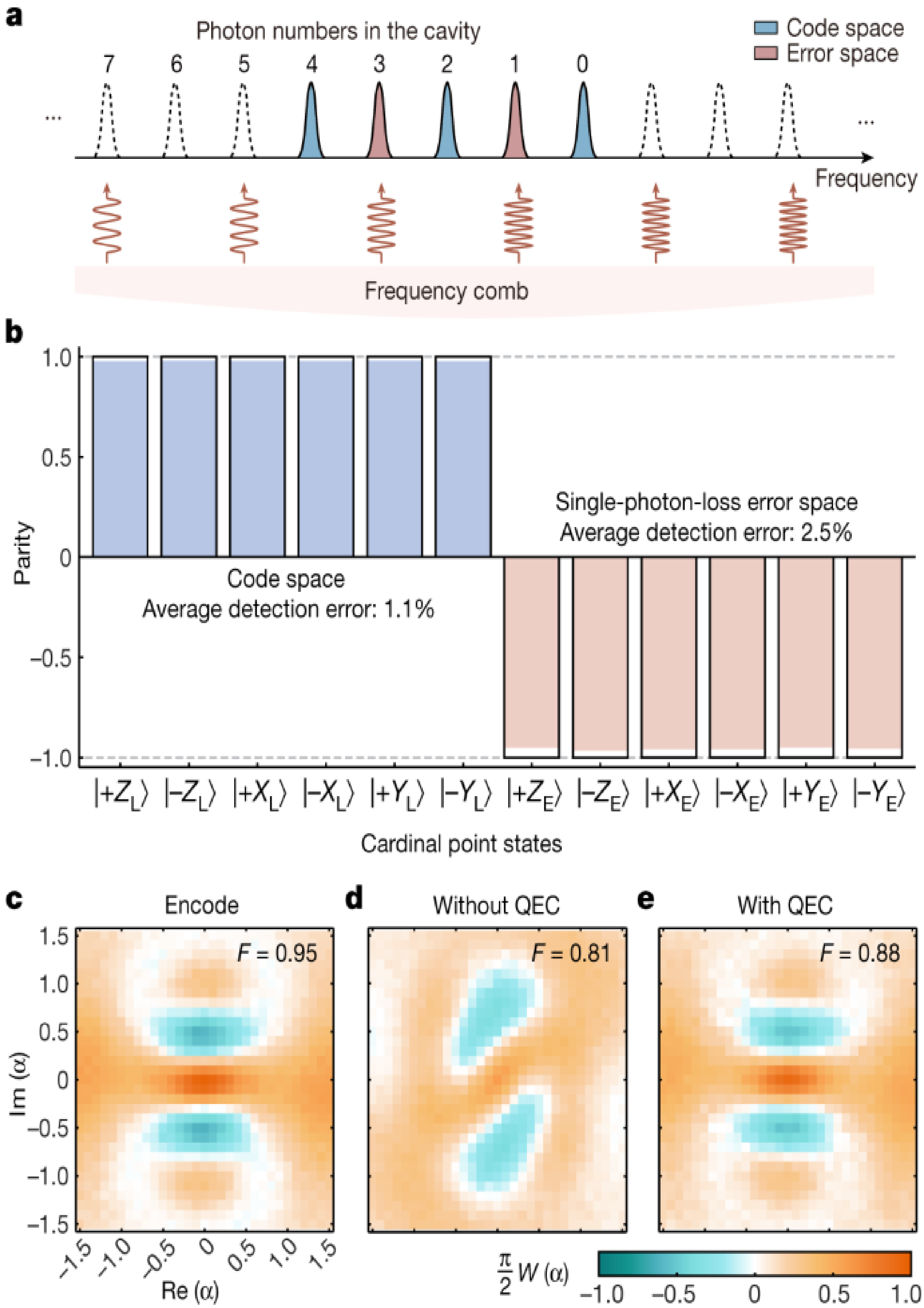
The auxiliary qubit is first encoded to the logical qubit in an oscillator with $\{\left|{0}_{{\rm{L}}}\right\rangle =\left(\left|0\right\rangle +\left|4\right\rangle \right)/\sqrt{2},\left|{1}_{{\rm{L}}}\right\rangle =\left|2\right\rangle \}$. Once a single-photon-jump error occurs, the logical qubit state falls out of the code space to the error space with the basis states: $\{\left|{0}_{{\rm{E}}}\right\rangle =\left|3\right\rangle ,\left|{1}_{{\rm{E}}}\right\rangle =\left|1\right\rangle \}$. After repetitive error detecting and correcting, the logical qubit state is protected against single-photon-jump errors. Finally, quantum state is decoded back to the auxiliary qubit for a final state characterization. The cardinal point states in the Bloch spheres of the code and error spaces are defined as $\left|+{Z}_{{\rm{L}}({\rm{E}})}\right\rangle =\left|{0}_{{\rm{L}}({\rm{E}})}\right\rangle ,\left|+{X}_{{\rm{L}}({\rm{E}})}\right\rangle =(\left|{0}_{{\rm{L}}({\rm{E}})}\right\rangle +\left|{1}_{{\rm{L}}({\rm{E}})}\right\rangle)/\sqrt{2}$ and $\left|+{Y}_{{\rm{L}}({\rm{E}})}\right\rangle =(\left|{0}_{{\rm{L}}({\rm{E}})}\right\rangle +i\left|{1}_{{\rm{L}}({\rm{E}})}\right\rangle)/\sqrt{2}$, respectively.

The experiments are performed with a circuit QED architecture[18], where a superconducting transmon qubit[38] as an auxiliary qubit is dispersively coupled to a three-dimensional microwave cavity[39,40,41]. The auxiliary qubit has an energy relaxation time of about 98 μs and a pure dephasing time of 968 μs, whereas the storage cavity has a single-photon lifetime of 578 μs (corresponding to a decay rate $\kappa_s/2\pi = 0.28$ kHz) and a pure dephasing time of 4.4 ms. The universal control of the multiple photon states of the cavity can be realized by using the anharmonicity of the auxiliary qubit, and thus the key stages of the QEC procedure, as illustrated in Fig. 1, can be achieved by encoding the logical qubit in the high-dimensional Fock spaces of the bosonic mode.

Our route towards the break-even points in the QEC is twofold: improving both the operation fidelity to the logical qubit and the error syndrome measurement fidelity. The first goal is achieved by using a tantalum transmon qubit with high coherence[42,43] and an optimal quantum control technique[44] with carefully calibrated system parameters (Methods). We attempt the second goal by an ingenious scheme of projection measurement of a selected collection of Fock states. The principle of the scheme is

illustrated in Fig. 2a, where a classical microwave pulse containing $2M$ frequency components is applied on the auxiliary qubit to read out the Fock states. Because the frequency of the auxiliary qubit depends on the photon number n (see Methods for more details), error syndrome detection is achieved by mapping the even parity to the auxiliary qubit ground state $\left|g\right\rangle$ (and the odd parity to the excited state $\left|e\right\rangle$) in a quantum non-demolition manner. This approach holds potential advantages of more flexible choices of error spaces and less sensitivity to auxiliary qubit damping and dephasing errors because the auxiliary qubit excitation is pronounced only when loss error occurs.

Fig. 2: Frequency comb control to measure the error syndrome.


a
Photon numbers in the cavity
7
6
5
4
3
2
1
0
Code space
Error space
Frequency
Frequency comb
b
Parity
1.0
0.5
0
–0.5
–1.0
Single-photon-loss error space
Average detection error: 2.5%
Code space
Average detection error: 1.1%
|+Z_L⟩ |–Z_L⟩ |+X_L⟩ |–X_L⟩ |+Y_L⟩ |–Y_L⟩ |+Z_E⟩ |–Z_E⟩ |+X_E⟩ |–X_E⟩ |+Y_E⟩ |–Y_E⟩
Cardinal point states
c
Encode
F = 0.95
d
Without QEC
F = 0.81
e
With QEC
F = 0.88
Im (α)
Re (α)
1.5
1.0
0.5
0
–0.5
–1.0
–1.5
π/2 W (α)
–1.0
–0.5
0
0.5
1.0

a, Frequency comb control is realized by mapping the photon number parity of the logical state to the auxiliary qubit state by applying a microwave pulse with multi-frequency components to the auxiliary qubit. Two components match the auxiliary qubit frequencies when the logical qubit is in the error space and other components are chosen symmetrically for the code space to eliminate the off-resonant driving effect on the logical states. **b**, Bar chart of the measured photon number parities for the six cardinal point states on the Bloch spheres of the logical qubit in the code and error spaces with the frequency comb parity measurement. Solid black frames correspond to the ideal parities ± 1 for the logical states in the code and error spaces. The numbers represent the average parity detection errors in these two spaces. **c**, Measured Wigner function of the cavity state after encoding the logical qubit in the $\left|+{X}_{{\rm{L}}}\right\rangle$ state. **d**,**e**, Measured Wigner functions of the same cavity state after a waiting time of about 90 μs without (**d**) and with (**e**) a single QEC operation. The numbers in these Wigner functions represent the corresponding state fidelities.

Source Data

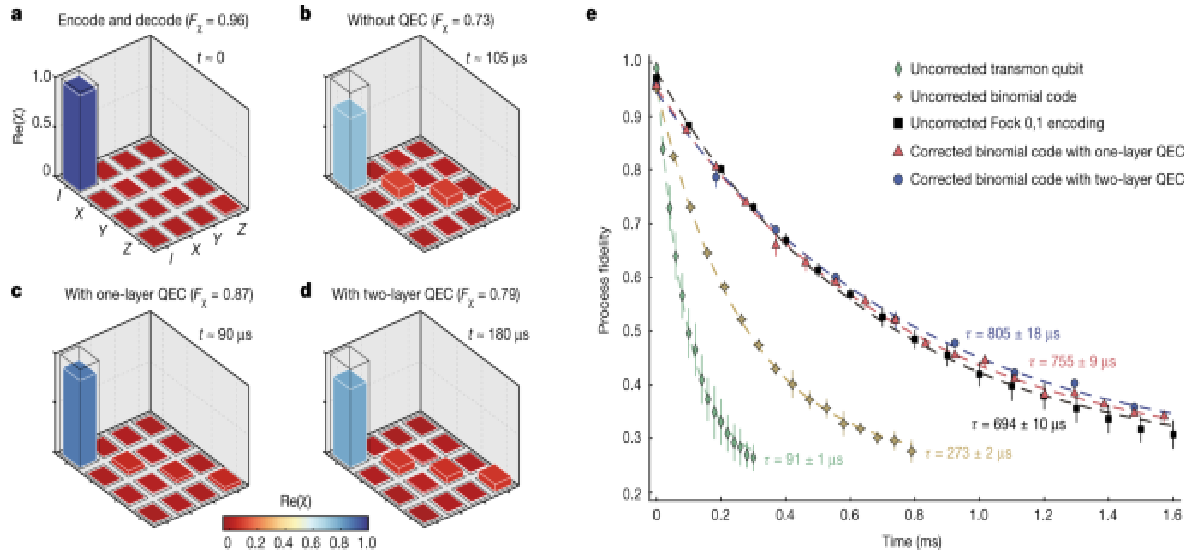
To characterize our syndrome measurement, the cavity is encoded to the six cardinal point states in the Bloch spheres of both the code and error spaces on the basis of the lowest-order binomial code words. The measured results of the cavity photon number parities are presented in Fig. 2b and show an average detection error of 1.1% and 2.5% for the cavity states in the code and error spaces, respectively. The encoding of the cavity, one of the most elementary processes of QEC, is further verified by the Wigner function with a high fidelity of 0.95, as shown in Fig. 2c.

On the basis of the above techniques, the QEC process of the binomial code can be implemented following the procedure in Fig. 1. However, practical imperfections limit the QEC performance: (1) during a waiting time of $t_{\rm w}$, that is, an idle process, there is a probability of about $2{({\kappa }_{{\rm{s}}}{t}_{{\rm{w}}})}^{2}\exp (-2{\kappa }_{{\rm{s}}}{t}_{{\rm{w}}})$ of a two-photon-loss error, which is undetectable for this lowest-order binomial code. (2) Owing to the non-commutativity of the single-photon-loss error and the self-Kerr interaction of the cavity, there is a large dephasing effect of the logical qubit induced by the unpredictable photon loss event, thus destroying the stored quantum information. (3)

Quantum recovery operations are imperfect. It is worth noting that there is a logical state distortion even if no photon loss is detected[8]. Strategies to mitigate the above imperfections are introduced, taking into account the whole system: choose an optimal waiting time, use a two-layer QEC procedure[17] to avoid unnecessary operation errors introduced by the error corrections and adopt the photon-number-resolved a.c. Stark shift (PASS) method[35] during idle operations to suppress photon-jump-error-induced decoherence in the code space (see Supplementary Information for more details). The measured Wigner functions of the cavity states after a single QEC cycle (about 90 μs of waiting) without and with performing the error correction operation are shown in Fig. 2d,e, with state fidelities of 0.81 and 0.88, respectively.

The performance of the QEC is benchmarked by the process fidelity \({F}_{\chi }\), which is defined as the trace of $\chi_{\rm exp}\chi_{\rm ideal}$, where $\chi_{\rm exp}$ denotes the experimentally measured process matrix for the QEC process and $\chi_{\rm ideal}$ is the ideal process matrix for an identity operation. In Fig. 3a, we present the measured process matrix for the encoding and decoding process only, which indicates a reference fidelity of 0.96. In the absence of a QEC operation after a waiting time of 105 μs, the process fidelity is reduced to a value of 0.73 because of the inability to protect the quantum information stored in the cavity from the single-photon-loss error, with the corresponding measured process matrix shown in Fig. 3b. When using the QEC operation, the process fidelity is improved because of the protection from the single-photon-loss error, with the process matrices for the one-layer and two-layer QECs shown in Fig. 3c,d, respectively.

Fig. 3: Performance of repetitive QEC operations.



a–**d**, Bar charts of the real parts of the process matrices for an encode and decode process (**a**), a waiting time of about 105 μs without QEC (**b**), a cycle time of about 90 μs with one-layer QEC operation (**c**) and a cycle time of about 180 μs with two-layer QEC operation (**d**). The numbers in brackets represent the process fidelities for each case. **e**, Process fidelity decays as a function of time for different encodings. Error bars correspond to 1 s.d. of several repeated measurements. The process fidelities for both the corrected binomial code with one-layer QEC (red triangles) and two-layer QEC (blue circles) exhibit slow decay, compared with the uncorrected Fock states $\{\left|0\right\rangle ,\left|1\right\rangle \}$ encoding (black squares), which defines the break-even point in this system. The corrected binomial code with two-layer QEC offers an improvement over the break-even point by a factor of 1.2, and also surpasses the uncorrected binomial code (yellow stars) by a factor of 2.9 and the uncorrected transmon qubit (green diamonds) by a factor of 8.8. All curves are fitted using $F_\chi = Ae^{-t/\tau} + 0.25$ to extract the lifetimes τ of the corresponding encodings. Uncertainties on τ are obtained from the fittings.

[Source Data](#)

The most important benchmark to characterize the performance of a QEC procedure is the gain in the lifetime of the protected logical qubit against that of the constituent element with the longest lifetime. For the three-dimensional circuit QED device, the best physical qubit is encoded with the two lowest photon-number states \(\{\left|0\right\rangle ,\left|1\right\rangle

\}\), which is more robust against decoherence effects than any other encoded photonic qubit without QEC protection. To quantitatively show the advantage of our QEC scheme, in Fig. 3e we present the measured process fidelities of the corrected binomial code as a function of the storage time with the repetitive one-layer (red triangles) and two-layer (blue circles) QECs, as well as those for the unprotected binomial code (yellow stars), the transmon qubit (green diamonds) and the Fock qubit (black squares) for comparison.

All curves are fitted according to the function $F_{\chi} = Ae^{-t/\tau} + 0.25$, with τ corresponding to the lifetime of the specific encoding and A being a fitting parameter. The offset in the fitting function is fixed to 0.25, implying a complete loss of information at the final time. As a result, the lifetime τ for the corrected binomial code with one-layer QEC is improved by about 8.3 times compared with the uncorrected transmon qubit and 2.8 times compared with the uncorrected binomial code. In particular, τ is improved to about 1.1 times that of the uncorrected Fock qubit encoding, that is, exceeding the break-even point of QEC in this system. Using the two-layer QEC scheme, the corresponding lifetime τ of the logical qubit is improved to about 8.8 times that of the uncorrected transmon qubit, 2.9 times that of the uncorrected binomial code and 1.2 times that of the break-even point. These results demonstrate that the quantum information stored in the cavity with multiphoton binomial encoding can indeed be preserved and protected from photon loss errors by means of repetitive QEC operations.

Table 1 shows an overall error analysis for the one-layer and two-layer QEC experiments. The error sources are divided into four parts: the intrinsic errors for the lowest-order binomial code, the error detection errors, the recovery operation errors and the auxiliary qubit thermal excitation errors during the QEC cycle. These errors can be estimated from either the numerical simulations or the measurement results of individual calibration experiments (Supplementary Information). The predicted lifetimes τ for the QEC experiments, calculated by \(\tau =-{T}_{{\rm{w}}}/ln(1-{\epsilon })\)[17], with T_{w} and ϵ being the total duration and the weighted total error per QEC cycle, are consistent with those in our QEC experiments.

Table 1 Error budget for the one-layer and two-layer QEC processes. The predicted lifetime calculated from the error model agrees well with the measured lifetimes

In conclusion, we experimentally demonstrate the prolonged coherence time of quantum information encoded with discrete variables in a bosonic mode by repetitive QEC. The break-even point has been reached by carefully designing the QEC procedure to balance the fidelity losses due to undetectable errors during the idle process, and error detection and correction operations. At present, the main infidelity is contributed by the two-photon-loss error that is beyond the ability of our current QEC code, but can be corrected by higher-order binomial codes[8]. Our frequency comb method could be used to measure the generalized photon number parity of such codes, enabling detection and correction of both single-photon-loss and two-photon-loss errors. Our work thus represents a key step towards scalable quantum computing and provides a practical guide for system optimization of quantum control and the design of the QEC procedure for future applications of logical qubits.

Methods

Experimental device and set-up

The circuit QED device in our experiment uses a hybrid three-dimensional–planar architecture[40] and consists of a superconducting transmon qubit[38], a coaxial stub cavity and a Purcell-filtered stripline readout resonator (see Fig. S1 in the Supplementary Information). The high-Q cavity is designed with a cylindrical re-entrant quarter-wave transmission line resonator[41], and machined from high-purity (99.9995%) aluminium. A horizontal tunnel is used to house a sapphire chip, on which the antenna pads of the transmon qubit and the striplines of the low-Q readout resonator are patterned with a thin tantalum film[42,43]. The single Al-AlO_x-Al trilayer Josephson junction of the transmon qubit is fabricated using a double-angle evaporation technique.

The fast feedback control is implemented with Zurich Instruments UHFQA and HDAWG, which are connected to each other through a digital input/output (DIO) link cable for real-time feedback control. The UHFQA

generates the readout pulses, acquires the down-converted transmitted readout signals for demodulation and discrimination in hardware, and sends the digitized readout results to the HDAWG through the DIO link cable in real time. The HDAWG plays different predefined waveforms conditional on the received readout results from the DIO link cable. The feedback latency, defined as the time interval between sending out the last point of the readout pulse from the UHFQA and sending out the first point of the control pulse from the HDAWG, is about 511 ns in our set-up, which also includes the time for the signal to travel through the experimental circuitry.

Parity mapping

The parity mapping procedure in the QEC experiment is implemented by applying a classical microwave pulse containing $2M$ ($M = 11$ in our experiment) frequency components on the auxiliary qubit, with the system dynamics governed by the Hamiltonian:

$$H/\hslash =-\chi {a}^{\dagger }a|e\rangle \,\langle e|+\Omega \,[\mathop{\sum }\limits_{n\,=\,1}^{2M}{e}^{-i{\delta }_{n}t}|e\rangle \,\langle \,g|+{\rm{h}}\,.{\rm{c}}\,.]$$
(1)

in the interaction picture. Here, $\left|e\right\rangle$ denotes the excited state and $\left|g\right\rangle$ denotes the ground state of the auxiliary qubit, $a^{\dagger}$ is the creation operator and a is the annihilation operator of the photonic field in the cavity, χ is the auxiliary qubit's frequency shift induced per photon as a result of its dispersive coupling, δ_n is the frequency detuning of the n-th driving component with a Rabi frequency of Ω and h.c. denotes the Hermitian conjugate. With the choice of the drive frequency detuning $\delta_n = (2M - 2n - 1)\chi$, the auxiliary qubit is resonantly driven when the cavity has $2m + 1$ photons with $m = 0, 1, \ldots M$.

For the cavity in the code space, the auxiliary qubit is off-resonantly driven by the comb pulse. For the two-photon state in the cavity, the qubit's transition $\left|g\right\rangle \leftrightarrow \left|e\right\rangle$ is driven by M pairs of frequency components with symmetric detunings, resulting in a qubit state revival at a time of $T = k\pi/\chi$ with k being an integer. Similarly,

for the zero-photon and four-photon states in the cavity, the qubit is driven by $M-1$ pairs of symmetric components and two unpaired components, whose effects can be ignored under the condition of $2M\chi \gg \Omega$. Therefore, the auxiliary qubit also makes a cyclic evolution at $T=k\pi/\chi$ and returns to the initial ground state when the cavity is in the code space.

For the cavity in the error space with one-photon and three-photon states, the auxiliary qubit's transition $\left|g\right\rangle \leftrightarrow \left|e\right\rangle$ is driven by a resonant frequency component, $M-1$ pairs of symmetric frequency components and an unpaired off-resonant component. Under the same condition of $2M\chi \gg \Omega$, we can neglect the off-resonant effect of the unpaired components, and the auxiliary qubit will evolve from the initial ground state to the excited state at $T=k\pi/\chi$, with k being an integer when choosing the drive amplitude $\Omega=\pi/2T$. In our experiment, $\Omega=\chi/4$, and $T\approx\pi/\chi$ for an optimized parity mapping time (see section II in the Supplementary Information).

Therefore, this frequency comb pulse achieves error syndrome detection by mapping the even parity of the cavity state to the auxiliary qubit $\left|g\right\rangle$ state (and the odd parity to the $\left|e\right\rangle$ state) in a quantum non-demolition manner. This parity mapping process can be intuitively illustrated by simultaneously applying two conditional π rotations to the auxiliary qubit to flip the qubit state to the excited state associated with the cavity's one-photon and three-photon states, thus resulting in a minimum perturbation to the cavity states in the code space.

Strategies for system optimization

The PASS method[35] is adopted to mitigate the photon-loss-induced dephasing effect of the logical code words, due to the non-commutativity of the annihilation operation and the self-Kerr term. In our experiment, we apply an off-resonant drive pulse with a frequency detuning of about -3.5χ on the auxiliary qubit during the idle operation, resulting in different phase accumulation rates f_n for Fock state $\left|n\right\rangle$ with $n=1, 2, 3, 4$ relative to the vacuum state. By choosing an optimal amplitude of the detuned drive, we could achieve the error-transparent condition[35] of (f_4-f_2)

$-(f_3 - f_1) = 0$ to mitigate the dephasing effect of the logical qubit (Supplementary Fig. 4).

To balance the operation errors, the no-parity-jump backaction errors and the photon-loss errors, we use a two-layer QEC procedure[17] to improve the QEC performance (see Fig. S6 in the Supplementary Information). In our QEC experiment, there are two bottom layers in a single QEC cycle: the first layer conserves the photon number parity in the deformed code space and the second layer recovers the quantum information in the code space.

The waiting time of the idle operation in each QEC cycle is selected on the basis of a trade-off between the uncorrected errors occurring during this time and the operation errors occurring during the error syndrome measurements and recovery operations. On the one hand, the longer the waiting time, the larger the probability of the two-photon-loss event occurring during this time, which cannot be detected by the lowest-order binomial code. On the other hand, the more frequent the error detection, the more likely it is that the photon-loss errors occur during the detections and corrections. We calculate the QEC lifetime as a function of the waiting time from numerical simulations and choose an optimal waiting time of about 90 μs in our QEC experiment (Supplementary Fig. 8).

Data availability

Source data for Figs. 2 and 3 are available with the paper. All other data relevant to this study are available from the corresponding author upon reasonable request.

Code availability

The code used for simulations is available from the corresponding author upon reasonable request.

References

1. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information*, 10th ed. (Cambridge Univ. Press, 2010).

2. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).

3. Calderbank, A. R. & Shor, P. W. Good quantum error-correcting codes exist. *Phys. Rev. A* **54**, 1098–1105 (1996).

4. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).

5. Cory, D. G. et al. Experimental quantum error correction. *Phys. Rev. Lett.* **81**, 2152–2155 (1998).

6. Knill, E., Laflamme, R., Martinez, R. & Negrevergne, C. Benchmarking quantum computers: the five-qubit error correcting code. *Phys. Rev. Lett.* **86**, 5811–5814 (2001).

7. Chiaverini, J. et al. Realization of quantum error correction. *Nature* **432**, 602–605 (2004).

8. Michael, M. H. et al. New class of quantum error-correcting codes for a bosonic mode. *Phys. Rev. X* **6**, 031006 (2016).

9. Terhal, B. M. Quantum error correction for quantum memories. *Rev. Mod. Phys.* **87**, 307–346 (2015).

10. Abobeih, M. H. et al. Fault-tolerant operation of a logical qubit in a diamond quantum processor. *Nature* **606**, 884–889 (2022).

11. Ryan-Anderson, C. et al. Realization of real-time fault-tolerant quantum error correction. *Phys. Rev. X* **11**, 041058 (2021).

12. Kelly, J. et al. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature* **519**, 66–69 (2015).

13. Chen, Z. et al. Exponential suppression of bit or phase errors with cyclic error correction. *Nature* **595**, 383–387 (2021).

14. Krinner, S. et al. Realizing repeated quantum error correction in a distance-three surface code. *Nature* **605**, 669–674 (2022).

15. Zhao, Y. et al. Realization of an error-correcting surface code with superconducting qubits. *Phys. Rev. Lett.* **129**, 030501 (2022).

16. Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature* **614**, 676–681 (2023).

17. Hu, L. et al. Quantum error correction and universal gate set operation on a binomial bosonic logical qubit. *Nat. Phys.* **15**, 503–508 (2019).

18. Blais, A., Grimsmo, A. L., Girvin, S. M. & Wallraff, A. Circuit quantum electrodynamics. *Rev. Mod. Phys.* **93**, 025005 (2021).

19. Campbell, E. T., Terhal, B. M. & Vuillot, C. Roads towards fault-tolerant universal quantum computation. *Nature* **549**, 172–179 (2017).

20. Waldherr, G. et al. Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).

21. Schindler, P. et al. Experimental repetitive quantum error correction. *Science* **332**, 1059–1061 (2011).

22. Egan, L. et al. Fault-tolerant control of an error-corrected qubit. *Nature* **598**, 281–286 (2021).

23. Postler, L. et al. Demonstration of fault-tolerant universal quantum gate operations. *Nature* **605**, 675–680 (2022).

24. Yao, X.-C. et al. Experimental demonstration of topological error correction. *Nature* **482**, 489–494 (2012).

25. Takeda, K., Noiri, A., Nakajima, T., Kobayashi, T. & Tarucha, S. Quantum error correction with silicon spin qubits. *Nature* **608**, 682–686 (2022).

26. Reed, M. D. et al. Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).

27. Córcoles, A. et al. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nat. Commun.* **6**, 6979 (2015).

28. Cai, W., Ma, Y., Wang, W., Zou, C.-L. & Sun, L. Bosonic quantum error correction codes in superconducting quantum circuits. *Fundam. Res.* **1**, 50–67 (2021).

29. Joshi, A., Noh, K. & Gao, Y. Y. Quantum information processing with bosonic qubits in circuit QED. *Quantum Sci. Technol.* **6**, 033001 (2021).

30. Flühmann, C. et al. Encoding a qubit in a trapped-ion mechanical oscillator. *Nature* **566**, 513–517 (2019).

31. Campagne-Ibarcq, P. et al. Quantum error correction of a qubit encoded in grid states of an oscillator. *Nature* **584**, 368–372 (2020).

32. Gertler, J. M. et al. Protecting a bosonic qubit with autonomous quantum error correction. *Nature* **590**, 243–248 (2021).

33. Ofek, N. et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature* **536**, 441–445 (2016).

34. Sivak, V. V. et al. Real-time quantum error correction beyond break-even. *Nature* https://doi.org/10.1038/s41586-023-05782-6 (2023).

35. Ma, Y. et al. Error-transparent operations on a logical qubit protected by quantum error correction. *Nat. Phys.* **16**, 827–831 (2020).

36. Reinhold, P. et al. Error-corrected gates on an encoded qubit. *Nat. Phys.* **16**, 822–826 (2020).

37. Chou, K. S. et al. Deterministic teleportation of a quantum gate between two logical qubits. *Nature* **561**, 368–373 (2018).

38. Koch, J. et al. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).

39. Paik, H. et al. Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).

40. Axline, C. et al. An architecture for integrating planar and 3D cQED devices. *Appl. Phys. Lett.* **109**, 042601 (2016).

41. Reagor, M. et al. Quantum memory with millisecond coherence in circuit QED. *Phys. Rev. B* **94**, 014506 (2016).

42. Place, A. P. M. et al. New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds. *Nat. Commun.* **12**, 1779 (2021).

43. Wang, C. et al. Towards practical quantum computers: transmon qubit with a lifetime approaching 0.5 milliseconds. *npj Quantum Inf.* **8**, 3 (2022).

44. Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T. & Glaser, S. J. Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172**, 296–305 (2005).

Acknowledgements

This work was supported by the Key-Area Research and Development Program of Guangdong Province (Grants No. 2018B030326001 and No. 2020B0303030001); the Shenzhen Science and Technology Program (Grant No. RCYX20210706092103021); the National Natural Science Foundation of China (Grants No. 12274198, No. 11904158, No. U1801661, No. 12274080, No. 12061131011, No. 92265210, No. 92165209, No. 11925404, No. 11890704 and No. 11875108); the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2022A1515010324); the Guangdong Provincial Key Laboratory (Grant No. 2019B121203002); the Program (Grant No. 2016ZT06D348); the Science, Technology and Innovation Commission of Shenzhen Municipality (Grant No. KYTDPT20181011104202253); the Shenzhen–Hong Kong cooperation zone for technology and innovation (Contract No. HZQB-KCZYB-

2020050); the National Key Research and Development Program of China (Grant No. 2017YFA0304303); the China Postdoctoral Science Foundation (BX2021167); the Innovation Program for Quantum Science and Technology (Grants No. ZD0301703 and No. ZD0102040201); and the Natural Science Foundation of Beijing (Grant No. Z190012).

Author information

Authors and Affiliations

1. Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, China

   Zhongchu Ni, Sai Li, Xiaowei Deng, Yanyan Cai, Libo Zhang, Fei Yan, Song Liu, Yuan Xu & Dapeng Yu

2. Guangdong Provincial Key Laboratory of Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, China

   Zhongchu Ni, Sai Li, Xiaowei Deng, Yanyan Cai, Libo Zhang, Fei Yan, Song Liu, Yuan Xu & Dapeng Yu

3. Department of Physics, Southern University of Science and Technology, Shenzhen, China

   Zhongchu Ni & Dapeng Yu

4. Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, China

   Weiting Wang & Luyan Sun

5. Fujian Key Laboratory of Quantum Information and Quantum Optics, College of Physics and Information Engineering, Fuzhou University, Fuzhou, China

   Zhen-Biao Yang & Shi-Biao Zheng

6. Beijing Academy of Quantum Information Sciences, Beijing, China

Haifeng Yu

7. International Quantum Academy, and Shenzhen Branch, Hefei National Laboratory, Shenzhen, China

Song Liu, Yuan Xu & Dapeng Yu

8. CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei, China

Chang-Ling Zou

9. Hefei National Laboratory, Hefei, China

Chang-Ling Zou & Luyan Sun

Contributions

Y.X. and D.Y. supervised the project. Y.X. conceived and designed the experiment. Z.N. performed the experiment. Z.N. and Y.X. analysed the data and carried out the numerical simulations. Z.N. and S. Li developed the feedback control technique under the supervision of Y.X., X.D., Y.C., W.W., Z.-B.Y. and F.Y. contributed to the experimental and theoretical optimization. L.Z., S. Liu and H.Y. provided supports in device fabrication. S.-B.Z. proposed the theoretical scheme of the frequency comb method. S.-B.Z., C.-L.Z. and L.S. provided theoretical and experimental supports. C.-L.Z., S.-B.Z., L.S. and Y.X. wrote the manuscript, and all authors provided feedback.

Corresponding authors

Correspondence to Luyan Sun, Shi-Biao Zheng, Yuan Xu or Dapeng Yu.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information

Supplementary Information

This file contains the following four sections and additional references: I. Experimental method; II. Frequency comb control method; III. Details of the QEC procedure; and IV. Error analysis.

Source data

Source Data Fig. 2

Source Data Fig. 3

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation,

distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Ni, Z., Li, S., Deng, X. *et al.* Beating the break-even point with a discrete-variable-encoded logical qubit. *Nature* **616**, 56–60 (2023). https://doi.org/10.1038/s41586-023-05784-4

- Received: 16 November 2022
- Accepted: 02 February 2023
- Published: 22 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05784-4

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Real-time quantum error correction beyond break-even**
    - V. V. Sivak
    - A. Eickbusch
    - M. H. Devoret

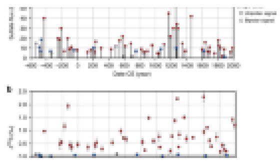
  Nature (2023)

Real-time quantum error correction beyond break-even

- V. V. Sivak
- A. Eickbusch
- M. H. Devoret

Nature Article 22 Mar 2023

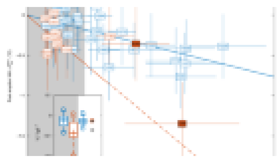
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05784-4



2600-years of stratospheric volcanism through sulfate isotopes

28 January 2019

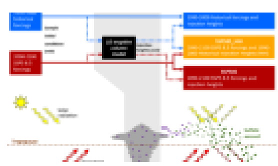
E. Gautier, J. Savarino, … J. Farquhar



Disproportionately strong climate forcing from extratropical explosive volcanic eruptions

28 January 2019

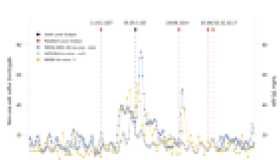
Matthew Toohey, Kirstin Krüger, … Rob Wilson



Climate change modulates the stratospheric volcanic sulfate aerosol lifecycle and radiative forcing from tropical eruptions

12 August 2021

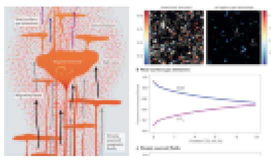
Thomas J. Aubry, John Staunton-Sykes, … Anja Schmidt



[Climatic and societal impacts of a “forgotten” cluster of volcanic eruptions in 1108-1110 CE]

21 April 2020

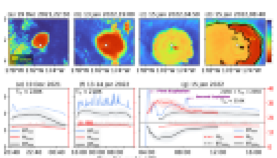
Sébastien Guillet, Christophe Corona, … Markus Stoffel



[The build-up and triggers of volcanic eruptions]

22 June 2021

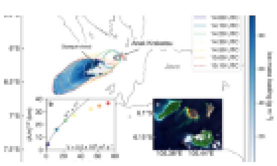
Luca Caricchi, Meredith Townsend, … Atsuko Namiki



[Eruption chronology of the December 2021 to January 2022 Hunga Tonga-Hunga Ha’apai eruption sequence]

13 December 2022

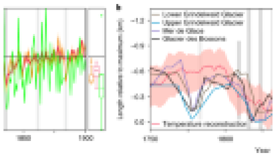
Ashok Kumar Gupta, Ralf Bennartz, … Tushar Mittal



[Anak Krakatau triggers volcanic freezer in the upper troposphere]

27 February 2020

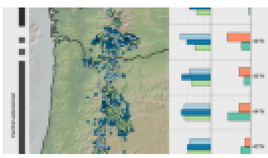
A. T. Prata, A. Folch, … A. Costa



[Last phase of the Little Ice Age forced by volcanic eruptions]

24 July 2019

Stefan Brönnimann, Jörg Franke, … Christoph C. Raible



[The causes of spatiotemporal variations in erupted fluxes and compositions along a volcanic arc]

22 March 2019

C. B. Till, A. J. R. Kent, … B. W. Pitcher

- Article
- Open Access
- Published: 05 April 2023

Lunar eclipses illuminate timing and climate impact of medieval volcanism

- Sébastien Guillet ,
- Christophe Corona,
- Clive Oppenheimer,

- Franck Lavigne,
- Myriam Khodri,
- Francis Ludlow,
- Michael Sigl,
- Matthew Toohey,
- Paul S. Atkins,
- Zhen Yang,
- Tomoko Muranaka,
- Nobuko Horikawa &
- …
- Markus Stoffel

Show authors

Nature **volume 616**, pages 90–95 (2023)

- 3729 Accesses
- 489 Altmetric
- Metrics details

Abstract

Explosive volcanism is a key contributor to climate variability on interannual to centennial timescales[1]. Understanding the far-field societal impacts of eruption-forced climatic changes requires firm event chronologies and reliable estimates of both the burden and altitude (that is, tropospheric versus stratospheric) of volcanic sulfate aerosol[2,3]. However, despite progress in ice-core dating, uncertainties remain in these key factors[4]. This particularly hinders investigation of the role of large, temporally clustered eruptions during the High Medieval Period (HMP, 1100–1300 ce), which have been implicated in the transition from the warm Medieval Climate Anomaly to the Little Ice Age[5]. Here we shed new light on explosive volcanism during the HMP, drawing on analysis of contemporary reports of total lunar eclipses, from which we derive a time series of stratospheric turbidity. By combining this new record with aerosol

model simulations and tree-ring-based climate proxies, we refine the estimated dates of five notable eruptions and associate each with stratospheric aerosol veils. Five further eruptions, including one responsible for high sulfur deposition over Greenland circa 1182 ce, affected only the troposphere and had muted climatic consequences. Our findings offer support for further investigation of the decadal-scale to centennial-scale climate response to volcanic eruptions.

Main

Large explosive volcanic eruptions can inject enormous quantities of sulfur-bearing gases into the stratosphere, where they generate sulfate aerosols[1]. The resulting aerosol veils perturb the energy budget of the Earth, inducing seasonal and regional surface temperature and precipitation anomalies, whose severity in combination with societal vulnerabilities has been linked to historical cases of agronomic and pasturage deficits, civil and political unrest, pestilence and migration[6]. Although the geologic record constitutes the primary evidence of past volcanism, with chronologies based on radiocarbon and other radiometric methods, polar ice cores arguably furnish the most comprehensive and accessible picture of climatically notable volcanism through the compilation of sulfur deposition time series[2,4]. Of particular note in such records is a proliferation of sulfur-rich eruptions during the HMP (circa twelfth and thirteenth centuries), beginning with a cluster of events around 1108–1110 ce (ref. [7]) and including the colossal Samalas eruption around 1257 ce (refs. [8,9]). These events have been linked with substantial cooling and subsistence crises[7,9] and the combined effect of their forcing has been posited as a contributor to the onset of the Little Ice Age[5].

The dating of past volcanic events from ice cores presents several challenges owing to the complexity of atmospheric transport leading to temporally and spatially variable sulfur deposition[10], poorly constrained age models[11,12,13] and uncertainties in layer counting related to accumulation rates and post-depositional processes[3]. A further challenge is the discrimination between tropospheric and stratospheric transport of volcanic aerosol, the latter being more indicative of a climate-forcing explosive eruption[4]. Sulfur isotopic

ratios measured in ice cores can help to make this distinction, but the approach has not been extensively applied and does not necessarily distinguish between tropospheric and lower stratospheric (below ozone layer) aerosol transport[3,14].

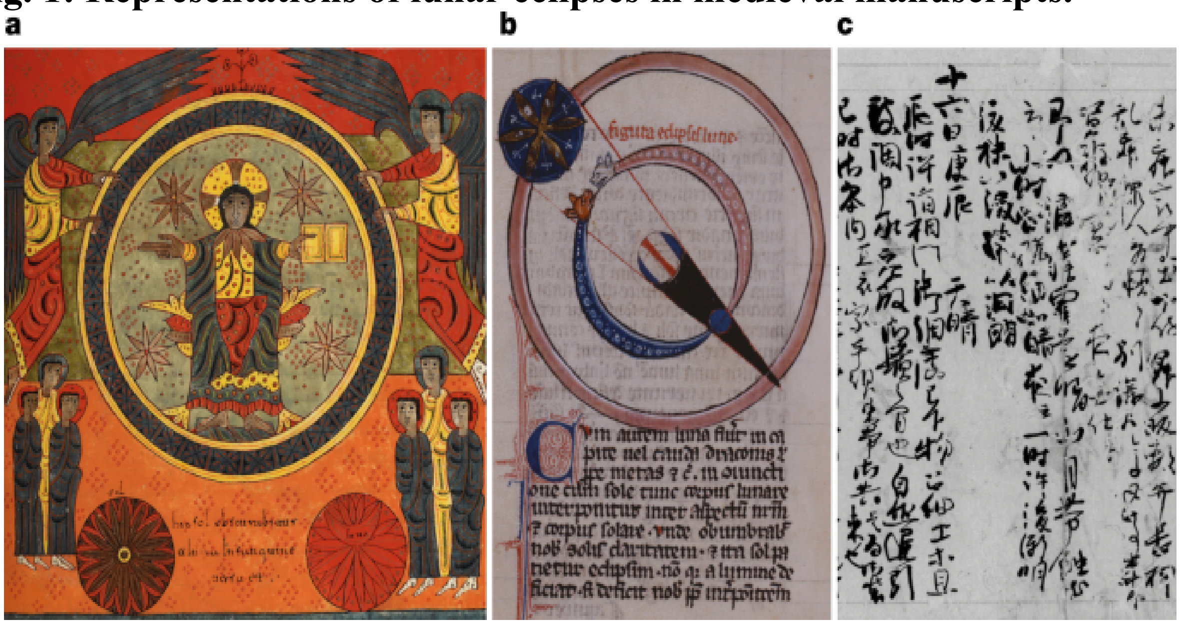
The rare and often visually spectacular atmospheric optical phenomena that can arise from the presence of volcanic dust veils in the stratosphere, such as solar dimming, coronae or Bishop's rings, peculiar twilight coloration and dark total lunar eclipses, have long been regarded as portents worth recording. References to such phenomena have provided independent evidence to evaluate the timing and impact of volcanism for the periods 1500 bce to 1000 ce (refs. [2,15]), 1500–1880 ce (refs. [16,17]) and 1880–2000 ce (refs. [18,19]). Here we focus on the notable lacuna in past studies, that is, the HMP, and on references in Eurasian sources to the coloration of total lunar eclipses, as they are relatively frequent and their occurrences are known precisely from astronomical retro-calculation. We derive an independent proxy for volcanic dust veil from medieval records of lunar eclipses and use the resulting time series, in conjunction with climate model outputs and summer temperature reconstructions from tree rings, to refine the NS1–2011 (Greenland) and WD2014 (Antarctica) ice-core chronologies, which hitherto have provided the primary constraints on timing of HMP eruptions[2,4]. The chronologies identify seven HMP eruptions that generated estimated volcanic stratospheric sulfur injections (VSSI) exceeding 10 Tg. Each of them ranks among the top 16 VSSI events of the past 2,500 years (refs. [2,4]). Their estimated eruption years are 1108 ce (UE1; in which UE stands for unidentified eruption; see Methods), 1171 ce (UE2), 1182 ce (UE3), 1230 ce (UE4), 1257 ce (Samalas), 1276 ce (UE5) and 1286 ce (UE6). We consider these events along with 13 lesser HMP eruptions and seek to confirm or refine the existing estimates of eruption year and season and to discriminate between tropospheric and stratospheric aerosol veils.

Eclipses unveil stratospheric turbidity

The brightness of the Moon during eclipse is highly sensitive to aerosol abundance in the stratosphere. Dark total lunar eclipses indicate high turbidity, whereas a ruddy disk signifies a clear stratosphere[18,19]. To reconstruct past stratospheric turbidity across the Medieval Climate

Anomaly to Little Ice Age transition, we comprehensively reexamined a large corpus of historical sources (Supplementary Dataset S1) written in the twelfth and thirteenth centuries, in search of credible lunar eclipse observations (Fig. 1). In China and Korea, lunar eclipses were recorded by official astronomers and preserved in sources such as the astronomical treatises and five-elements treatises of official dynastic histories, whereas Japanese eclipse observations are found in more diverse writings such as the diaries of courtiers, chronicles or temple records. In Europe, annals and chronicles from monasteries and towns represent the main sources. In Arabic sources, lunar eclipse observations are frequently found in universal chronicles[20].

Fig. 1: Representations of lunar eclipses in medieval manuscripts.



a, *Commentary on the Apocalypse* by Beatus of Liébana, from the monastery of Santo Domingo de Silos, near Burgos, Spain, 1090–1109 ce. Credit: British Library Board (Add. MS 11695, f108r). The text at the bottom of the miniature, between the dark circle on the left representing a total solar eclipse and the red circle on the right representing a total lunar eclipse, reads: “hic sol obscurabitur et luna in sanguine versa est” (“and the Sun was obscured and the Moon turned into blood”). The blood-red eclipsed Moon was seen as one possible sign of the Apocalypse. Lunar occultation descriptions from the Middle Ages often follow the Book of Revelation, suggesting that the Bible provided justification and inspiration for recording

lunar eclipses and their colour. **b**, Thirteenth-century depiction of a lunar eclipse by Johannes de Sacrobosco. Credit: The New York Public Library (*De Sphaera*, MssCol 2557, f112v). **c**, Facsimile of the *Meigetsuki* (明月記) diary by Fujiwara no Teika (藤原定家) describing the total lunar eclipse of 2 December 1229 ce. Credit: *Meigetsuki*, vol. 4, pp. 517, 2000. Reizei-ke Shiguretei Bunko. Tokyo: Asahi Shinbunsha. Teika mentions this event twice. The figure shows the first entry: "[…] the sky was free of clouds into the distance and the Moon over the hills emerged in eclipse, total for a little while, [its light] meagre as on a dark night. About an hour later it brightened gradually, and after it was extinguished [during the eclipse] it was especially luminous". The second entry, written four days later, details the unusual coloration of the Moon. Over the centuries, several portions of the Meigetsuki were cut apart, and the entry for 6 December 1229 ce is held in a private collection[49,50].

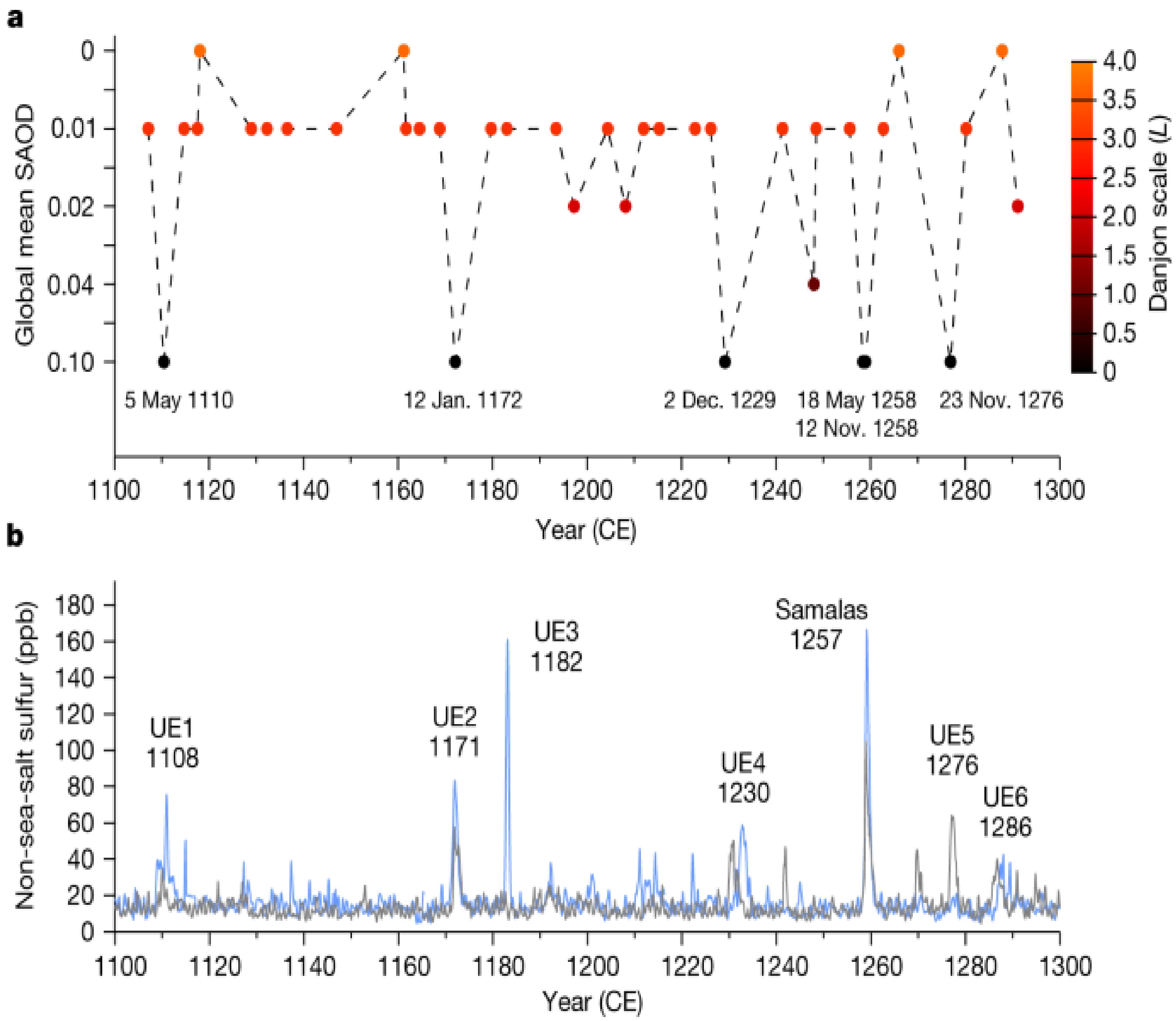
According to the latest catalogues of lunar eclipses[21,22], 64 (Europe), 59 (Middle East) and 64 (East Asia) total lunar eclipses occurred and would have been visible, weather permitting, between 1100 and 1300 ce. A total of 180 European, 10 Middle Eastern and 199 East Asian accounts describe 51, 7 and 61 individual total lunar eclipses, respectively. In Europe, although 12 individual eclipses are described in only one surviving source, many are corroborated in several accounts, up to 16 in the case of the eclipse of 11 February 1161 ce (Supplementary Dataset S1). This success in finding observations of retro-calculated eclipses for Europe (80%; Extended Data Table 1) is notable and comparable with that of later periods for which documentation is more abundant[16,17] (for example, 82%, 1665–1881 ce). It reflects the contemporary proliferation and geographic span of monastic communities across Europe[20], which improved the overall chances of clear-sky observations, and the attention that some chroniclers paid to celestial phenomena[23]. Observations of the Sun, the Moon or the stars were required to calculate the hours for prayer, as not all monasteries possessed water clocks or astrolabes for timekeeping[24,25,26]. Accurate observations of the age of the Moon were also important for the correct identification of the Easter full Moon, which served as the key point of reference for Easter Sunday and all other moveable feast days of the liturgical year[25,27]. East Asian records sometimes contain predictions rather than observations[28] and, for this

reason, the proportion of eclipses documented by observers in China, Korea and Japan was not computed.

Dark eclipses follow substantial HMP eruptions

Western and Eastern Christian sources together provide information on the colour and brightness of the Moon for 36 eclipses (Fig. 2a). Such attention to brightness is largely absent for Asian records[20,29], in which only one account describes coloration. References to a ‘blood-red Moon’ in Western and Eastern Christian sources are probably informed by texts such as the Book of Revelation of John, in which the blood Moon, along with earthquakes and solar eclipses, portended the End Times (Revelation 6:12–17; Fig. 1a). Lunar eclipse coloration was imbued with particular significance for Christian observers and often regarded as an ill omen, presaging disasters[26,30,31], emphasizing the influence of the Bible on the perception of natural phenomena during the Middle Ages[32]. This does not mean that the physical causes of lunar eclipses were unknown to all medieval chroniclers[31]. Ancient Babylonian, Greek and later Muslim astronomers not only understood eclipse mechanisms but were able to predict lunar occultations[20], and this knowledge was ultimately transmitted to medieval Europe, as evident in contemporary astronomical treatises (for example, *De Lunationibus*, 1112 ce, and *De Dracone*, 1120–1121 ce, by Walcher of Malvern[33]; *De Sphaera*, circa 1230 ce, by Johannes de Sacrobosco[34]; Fig. 1b). Both natural and supernatural interpretations of lunar eclipses coexisted in the Middle Ages[31], underpinning the recovery of a near-complete series of lunar eclipse coloration spanning two centuries (Fig. 2a).

Fig. 2: Stratospheric turbidity derived from total lunar eclipse coloration and non-sea-salt sulfur records from polar ice cores.



a, Total lunar eclipse descriptions retrieved from European, Middle Eastern and East Asian historical sources from 1100 to 1300 ce (Supplementary Dataset S1), rated on the Danjon scale (right *y* axis), and converted to equivalent global mean $SAOD_{550}$ (SAOD at 550 nm; left *y* axis). **b**, Monthly resolved non-sea-salt sulfur concentrations from the Greenland NEEM-2011-S1 (blue line) and Antarctica WDC06A (grey line) ice cores[2].

Source data

We rated the colour and luminosity of each observed eclipse on the Danjon scale[35], which quantifies naked-eye lunar brightness. It ranges from $L = 0$ (very dark) to $L = 4$ (very bright copper-red or orange eclipse). Of the 37 total lunar eclipses with brightness recorded in Eurasian sources, only six were rated $L = 0$, emphasizing the rarity and significance of such

observations (Fig. 2a and Supplementary Dataset S1). These events occurred on the night of 5–6 May 1110 ce, 12–13 January 1172 ce, 2–3 December 1229 ce, 18–19 May 1258 ce, 12–13 November 1258 ce and 22–23 November 1276 ce. All testimonies are reported in Supplementary Dataset S1, with each description stressing a near-complete and prolonged disappearance of the lunar disk. One of the most outstanding accounts was retrieved from Japanese sources and pertains to the total lunar eclipse of 2 December 1229 ce. Although Asian sources rarely detail coloration[20,29], the *Meigetsuki* (明月記, *The Record of the Bright Moon*) written by Fujiwara no Teika (藤原定家, 1162–1241 ce) reports an extremely dark lunar eclipse despite clear weather. The *Meigetsuki* notes that the coloration of the Moon was deemed so unusual that the astronomers expressed fear over its appearance: "Regarding the recent total lunar eclipse, although on previous occasions there has been totality, the old folk had never seen it like this time, with the location of the disk of the Moon not visible, just as if it had disappeared during the eclipse. Moreover, the duration was very long, and the change was extreme. It was truly something to fear. Indeed, in my seventy years I have never heard of or seen [such a thing]; the official astronomers spoke of it fearfully […]" (Fig. 1c and Supplementary Dataset S1).

All the dark ($L = 0$) lunar eclipses—in May 1110 ce, January 1172 ce, December 1229 ce, May 1258 ce, November 1258 ce and November 1276 ce—are contemporary with five of the seven largest HMP volcanic sulfate signals recorded in polar ice cores (UE1, UE2, UE4, Samalas and UE5; Fig. 2b), strongly suggesting that the darkening of the eclipsed Moon was related to the presence of volcanic aerosols in the stratosphere. This finding mirrors previous work that found all very dark total lunar eclipses since 1600 ce followed substantial volcanic eruptions[16,17,18,19,36] (Extended Data Table 2). For the remaining two of the top seven HMP eruptions, circa 1182 (UE3) and 1286 ce (UE6), descriptions of reddish ($L = 3–4$) total lunar eclipses point to low stratospheric turbidity in August 1179, December 1182 and October 1287 ce.

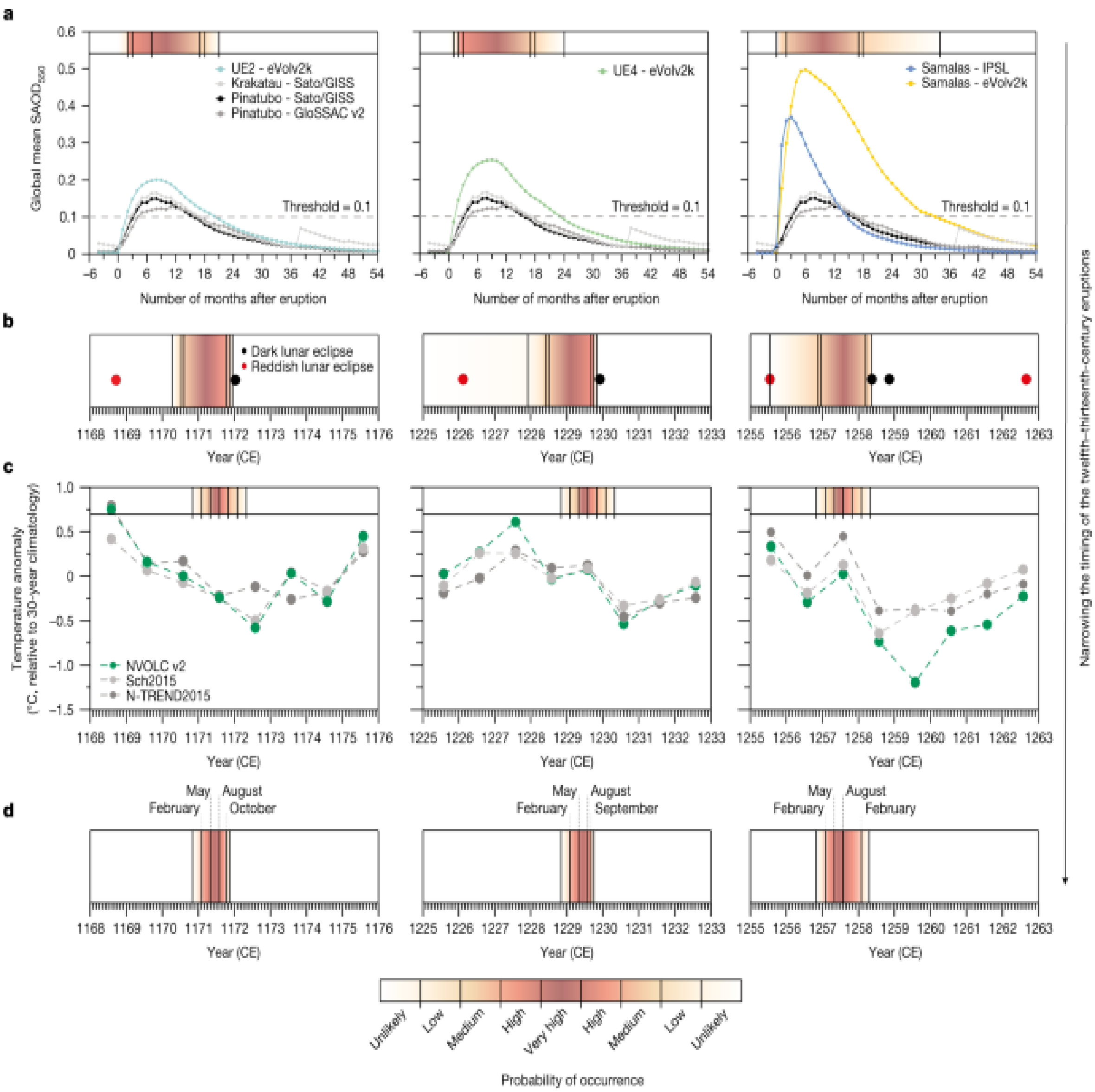
Timing of HMP eruptions

We constrained the timing of HMP eruptions by developing a four-step procedure that integrates evidence from our eclipse record, global aerosol simulations, modern satellite observations and tree-ring reconstructions (see Methods and Extended Data Fig. 1). First, the appearance of lunar eclipses reported in historical archives was rated on the Danjon scale and converted to stratospheric aerosol optical depth (SAOD) following refs. [16,17,18,19], which showed that, for 46 well-observed lunar eclipses between 1880–1888 ce and 1960–2001 ce, dark total lunar eclipses ($L = 0$) only occurred when SAOD exceeded about 0.1. Next, drawing on observations for the 1883 Krakatau and 1991 Pinatubo eruptions (from the Sato/GISS and GloSSAC v2 datasets)[37,38], SAOD simulations[39] (from the eVolv2k datasets for UE1 to UE6 and both the circa 1257 ce Samalas and 1815 ce Tambora eruptions) and IPSL-CM5A-LR[40] climate model outputs (for Samalas and Tambora), we evaluated the post-eruptive duration of elevated stratospheric turbidity, that is, SAOD ≥ 0.1. This suggested that a total lunar eclipse is most probably observed as dark between 3 and 20 months following an eruption. Thus, we assume that, given an observation of a dark lunar eclipse, the eruption responsible occurred between 20 and 3 months beforehand. This assumption is corroborated if we consider the seven largest VSSI events since 1600, which were all followed by dark lunar eclipses between 9 (1912 ce Katmai, 1982 ce El Chichón), 14 (1815 ce Tambora, 1883 ce Krakatau), 18 (1991 ce Pinatubo) and 20 (1600 ce Huaynaputina) months later (Extended Data Table 2). The case of the 1963 ce Agung eruption provides further corroboration of our approach because, of the three lunar eclipses that occurred 10, 15 and 21 months after this eruption, only the last (falling outside our 3–20-month window) was not rated $L = 0$ on the Danjon scale (Extended Data Table 2). Then, to provide further constraint on the timing of each HMP eruption, we evaluated independently dated Northern Hemisphere tree-ring-based summer temperature reconstructions for post-volcanic climatic response (Sch2015 (ref. [41]), N-TREND2015 (ref. [42]), NVOLC v2 (ref. [7])).

Taking UE2 (Fig. 3a) as an example, combining the dark lunar eclipse date of January 1172 ce with the eVolv2k, Sato/GISS and GloSSAC v2 datasets, we find a high to very high probability that the event occurred between July 1170 and October 1171 ce (Fig. 3b). The peak cooling observed in Sch2015 and NVOLC v2 in summer 1172 ce reduces the

likelihood of an eruption occurring between summer 1170 and autumn 1171 ce (Fig. 3c); we thus bracket the eruption date between May and August 1171 ce (Fig. 3d). We similarly constrain the probable time windows of other large HMP events to Northern Hemisphere winter 1108/1109 ce (UE1; Extended Data Fig. 2), Northern Hemisphere spring/summer 1229 ce (UE4) and Northern Hemisphere spring/summer 1257 ce for Samalas (Fig. 3d). This refined timing for the Samalas eruption is consistent with the isopach pattern of tephra fall in Indonesia[8], pointing to a dry-season eruption (between May and October) and counters an argument for a 1256 ce eruption date[43]. For UE5, we find a time window between September 1275 and July 1276 ce (Extended Data Fig. 2). Further constraint is not possible in this case as pronounced summer cooling is not evident in the tree-ring-based temperature reconstructions (Extended Data Fig. 2).

Fig. 3: Constraining the timing of HMP (1100–1300 ce) volcanic eruptions.



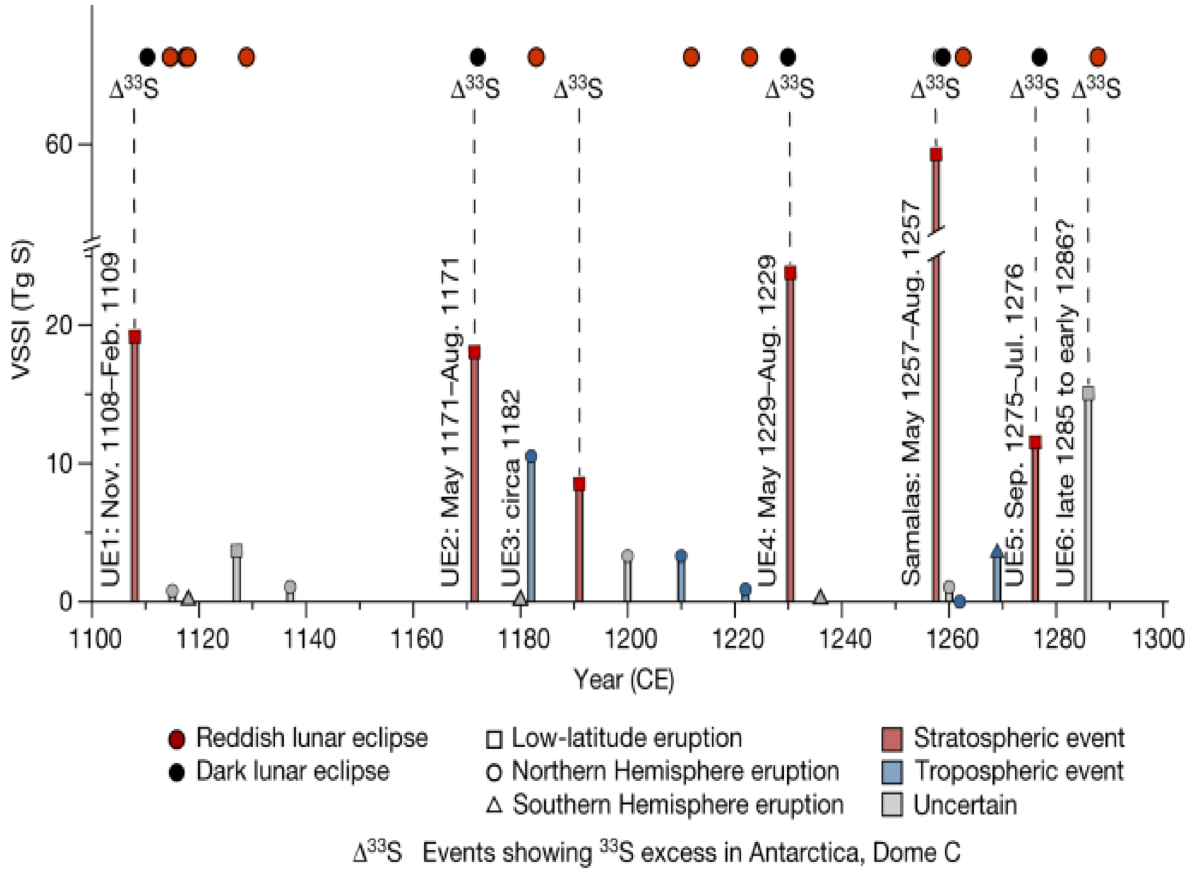
a, Residence time of volcanic stratospheric aerosols and time windows with SAOD exceeding about 0.1. The residence time of aerosols is based on global mean $SAOD_{550}$ time series from the Sato/GISS[37] and GloSSAC v2 (ref. [38]) datasets (for the 1883 ce Krakatau and 1991 ce Pinatubo eruptions) and simulated by the EVA forcing generator[4,39] (for UE2, UE4 and the 1257 ce Samalas eruption) and the IPSL-CM5A-LR model[40] (for the Samalas eruption). Probability of occurrence of HMP eruptions based on the timing of dark lunar eclipse dates (**b**) and tree-ring records[7,41,42] (**c**). **d**, Integration of **b** and **c** to estimate the most probable time windows for UE2, UE4 and the Samalas eruption.

Source data

Stratospheric or tropospheric aerosols?

Stratospheric aerosols profoundly affect the brightness and coloration of the eclipsed Moon by reducing the transmission of sunlight into the Earth's umbra[18,19]. Here we use lunar eclipse coloration to distinguish between tropospheric and stratospheric aerosol veils (see Methods). The dark lunar eclipses observed after UE1, UE2, UE4, Samalas and UE5 indicate stratospheric aerosol veils (Fig. 4), consistent with $\Delta^{33}S$ isotope records from Dome C, Antarctica[3]. Although UE3 (circa 1182 ce) registers as the second greatest Northern Hemispheric extratropical eruption of the last millennium in terms of ice-core sulfate deposition[4] (second only to Laki 1783–1784 ce), the reddish lunar eclipse of 11 December 1182 ce (Fig. 2a) points to a comparatively low stratospheric aerosol burden, while tree-ring reconstructions show limited cooling (Supplementary Dataset S2). Further consideration of eclipse coloration, in conjunction with tree-ring-based summer temperature reconstructions, suggests that for four HMP eruptions associated with lesser VSSI around 1127, 1210, 1222 and 1262 ce, sulfate aerosols were mostly confined to the troposphere and any climatic impacts were limited (Fig. 4 and Supplementary Dataset S2).

Fig. 4: Revised chronology of explosive volcanism in the twelfth and thirteenth centuries.



Vertical bars are based on the eVolv2k volcanic forcing reconstruction[4] and indicate the magnitude of VSSI. Using total lunar eclipse coloration (red and black dots) and $\Delta^{33}S$ isotope records[3], we discriminate between stratospheric (red bars) and tropospheric (blue bars) dust veils. Grey bars show uncertain events. Squares, circles and triangles refer to low-latitude, Northern Hemisphere extratopical and Southern Hemisphere extratropical eruptions, respectively[2].

Source data

The nature and timing of UE6 remain enigmatic (Fig. 4). A bipolar sulfate deposition distribution is observed around 1286 ce, implying a tropical eruption, whereas sulfur isotopic analysis suggests that the associated Antarctic deposition was stratospheric in origin[20]. However, the reddish lunar eclipse of 22 October 1287 ce (Fig. 2a), described in three independent records (Supplementary Dataset S1) from England and Italy, precludes substantial Northern Hemisphere stratospheric aerosol presence at this time. This apparent discrepancy may reflect a late 1285 or early 1286 ce eruption,

thus being too early to darken the October 1287 ce lunar eclipse. Regardless, the tree-ring proxies do not show substantial Northern Hemisphere cooling during the period 1280–1290 ce, suggesting limited climatic impacts of UE6 (Supplementary Dataset S2).

Implications

Our identification of marked stratospheric dust veils in 1110, 1172, 1229, 1258 and 1276 ce using contemporary observations of dark total lunar eclipses corroborates the general accuracy of the revised ice-core chronologies for Greenland (NS1–2011) and Antarctica (WD2014), while adding precision to the chronological framework of HMP eruptions. Given the inherent uncertainties in ice-core chronologies, our dataset of precisely dated dark total lunar eclipses offers a new, reliable and independent suite of chronological tie points that can complement established age markers in 536, 774/5, 939, 993/4, 1258, 1601 and 1816 ce to aid future chronology development. Our findings also suggest that five other events, likely associated with tropospheric-only aerosol veils, had a limited impact on climate.

No single source or method can, however, offer complete chronological control, and our eclipse data also have limitations (see Methods). The visibility of lunar eclipses is geographically and meteorologically constrained; accounts of partial and penumbral eclipses cannot be used, reducing the number of available records; and only comments on the colour of the Moon are relevant. We have thus developed here a multiproxy approach using diverse sources and methods that reflect or model different aspects of the volcano–climate system, each offering complementary constraints on eruption timing. These include ice-core sulfate deposition profiles, aerosol model simulations and palaeoclimatic proxies, as well as our eclipse observations. Further developing such integrative approaches will pave the way to even finer temporal resolutions, in particular as the representation of stratospheric aerosol formation, evolution and duration in climate models and the resolution of palaeoclimate records continue to improve (ref. [44]; see Methods).

Better constraints (ideally sub-annual) on the timing of historical explosive volcanic eruptions are critical because climatically important factors including aerosol distribution, altitude, size and radiative forcing are all influenced by the seasonally changing stratospheric circulation[45,46]. For unidentified eruptions, climate modelling typically uses notional eruption dates (for example, 1 April in the Community Earth System Model, 1 January in eVolv2k)[39,47] and assumes stratospheric aerosol presence, biasing the distribution, magnitude and persistence of modelled post-eruption thermal and hydroclimatic anomalies and potentially contributing to persistent model–proxy discrepancies[45,48]. Our findings thus offer improved parameterizations for the next generation of Community Earth System Models in investigations of the impacts of HMP eruptions. We hope that our new dataset will help to inform the extent of their role in the onset of the Little Ice Age.

Methods

Description of the twelfth–thirteenth-century eruptions investigated

Ice-core records from Greenland and Antarctica suggest the occurrence of nine low-latitude eruptions dated, according to the NS1–2011 and WD2014 timescales[2,4], to 1108, 1127, 1171, 1191, 1230, 1257, 1260, 1276 and 1286 ce, seven Northern Hemisphere extratropical events (1115, 1137, 1182, 1200, 1210, 1222 and 1262 ce, identified by deposition signals in Greenland only) and four Southern Hemisphere extratropical events (1118, 1180, 1236 and 1269 ce, identified by deposition signals in Antarctica only) in the twelfth and thirteenth centuries. This period therefore represents one of the most volcanically perturbed periods of the past 2,500 years. The largest volcanic eruptions—with VSSI > 10 Tg S in ref. [4] —are UE1 (1108 ce), UE2 (1171 ce), UE3 (1182 ce), UE4 (1230 ce), the Samalas (circa 1257 ce) eruption[8], UE5 (1276 ce), UE6 (1286 ce) and rank as the 7th (VSSI, 19.2 Tg S), 10th (18.1 Tg S), 16th (10.1 Tg S), 4th (23.8 Tg S), 1st (59.4 Tg S), 15th (11.5 Tg S) and 13th (15.1 Tg S) largest volcanic events of the last millennium by sulfate deposition. With the exception of the circa 1257 ce event, attributed to Samalas in Indonesia[8], the sources of these eruptions

remain unidentified. Although the 1108 ce sulfate spike was originally attributed to an eruption of a tropical volcano[2], a recent reassessment of ice-core records coupled with historical sources suggests that at least two eruptions occurring between 1108 and 1110 ce are registered in the observed polar sulfate deposition, one in the Northern Hemisphere extratropics and one in the tropics[7].

Investigation of historical sources and development of a lunar eclipse database

European and Middle Eastern lunar eclipse observations

We extensively reexamined hundreds of annals and chronicles, written in the twelfth and thirteenth centuries ce, searching for references to lunar eclipses. For Europe, contemporary sources are mostly accessible in compilations of medieval texts edited in the series of the (1) *Monumenta Germaniae Historica*, (2) *Rerum Britannicarum Medii Ævi*, (3) *Recueil des historiens des Gaules et de la France* and (4) *Rerum Italicarum Scriptores*. Descriptions or observations of lunar eclipses originate from across Europe, namely, Austria, the Czech Republic, England, France, Germany, Iceland, Italy, Sweden and Switzerland. In a few instances, lunar obscurations were observed in the Middle East by Christians engaged in the Crusades. These chronicles were found in the *Recueil des historiens des croisades*, edited and published in the nineteenth century. Most sources consulted were composed by monks or clerics and, to a lesser extent, by urban laymen. Nearly all were composed in Latin, although the vernacular was occasionally used. Furthermore, we included observations retrieved from Ukrainian and Russian chronicles in our database[51]. We also examined observations of lunar eclipses recorded in Arabic chronicles based on an existing compilation[52]. However, because none of these sources contains information on the colour of the eclipsed Moon, we did not analyse them further.

East Asian lunar eclipse observations

In China and Korea, eclipse records are mainly found in the *Treatise on Astronomy*, *Treatise on the Calendar* and *Treatise on the Five Elements* of

the official dynastic histories[20]. For China, we reexamined the *Song Shi* (宋史), the official history of the Song dynasty (960 to 1279 ce); the *Jin Shi* (金史), the official history of the Jin dynasty (1115–1234 ce); and the *Yuan Shi* (元史) the official history of the Yuan dynasty (1261 to 1367 ce). For Korea, we investigated the *Goryeosa* (高麗史), the history of the Goryeo dynasty (936 to 1392 ce). In Japan, the astronomical records are found in a variety of works ranging from privately and officially compiled histories to diaries of courtiers and temple records[20]. We thus focused on the lunar eclipse observations compiled in the benchmark work, the *Nihon Tenmon Shiryô* (日本天文史料)[53], by Japanese astronomer S. Kanda. This work lists solar and lunar eclipses, comets and aurorae and provides original texts. The most pertinent descriptions of lunar eclipses derive from the *Meigetsuki* (明月記, 1180–1235 ce; the diary of Fujiwara no Teika) and the *Azuma Kagami* (吾妻鏡, 1180–1266 ce; *Mirror of the East*, a chronology of the Kamakura shogunate). Chinese sources are written in classical Chinese, whereas Japanese and Korean sources are written in Sino-Japanese (Kanbun) and Sino-Korean (Hanmun), respectively. However, East Asian sources rarely report information on the colour of total lunar eclipses; only the eclipse of 2 December 1229 ce could be used to estimate stratospheric turbidity.

Assessing the reliability of historical sources

The reliability of each eclipse observation was assessed by historical source criticism and by reference to the five-millennium catalogue of lunar eclipses (1999 bce to 3000 ce)[21] and the eight-millennium catalogue of lunar eclipses (4000 bce to 4000 ce)[22], which uses the most up-to-date historical values of Delta (ΔT)[54,55]. We also ensured that all the eclipses reported in historical sources were visible (where relevant) in Europe, the Middle East and East Asia using the visibility maps and the local circumstances tables provided by both catalogues. Care was taken to identify second-hand reports, that is, those that a given author did not witness but instead drew information on from another source. Frequent duplication occurred in Western and Eastern Christian sources owing to common underlying source materials and the scribal practices of copying, synthesizing and piecemeal updating of annals and chronicles. A table listing all total lunar eclipses records for

which information about the colour of the Moon could be retrieved, and providing summary context for the historical sources investigated, is presented in Supplementary Dataset S1. An extended version of Supplementary Dataset S1 containing all descriptions of total lunar eclipses from Europe, the Middle East and East Asia considered in this study (with or without reference to colour), along with the eclipse visibility maps, can be accessed from the Zenodo repository: https://doi.org/10.5281/zenodo.6907654.

Four-step procedure to refine the dating of HMP eruptions

To contrain the timing of HMP eruptions, we developed a four-step methodology based on analysis of historical sources (step 1); global aerosol simulations and observations (step 2); tree-ring-based temperature reconstructions (step 3); and integration of results of steps 1–3 (step 4). A more detailed breakdown of each step and a specific illustration of the procedure using the well-known example of the 1815 ce Tambora eruption are provided below.

Deriving SAOD from accounts of total lunar eclipse coloration and brightness

During a total lunar eclipse, as the Moon passes through the shadow of the Earth, it is partially illuminated by light refracted by the atmospheric limb. The spectrum of the refracted sunlight is influenced by scattering and absorption. Rayleigh (molecular) scattering is stronger at shorter wavelengths, least affecting orange or red-coloured light. When the stratosphere is little perturbed, the eclipsed Moon thus tends to appear copper to deep red. With a turbid stratosphere, scattering of visible light is strongly enhanced at all wavelengths, diminishing transmission through the atmospheric limb such that the Moon in eclipse appears dark(er). In extreme cases, it can appear to vanish almost completely[16,17,18,36,56,57]. The colour and luminosity (L) of the eclipsed Moon were rated according to the Danjon scale[35], which runs from $L = 0$ to $L = 4$:

- $L = 0$: very dark eclipse. Moon almost invisible, especially at mid-totality.

- *L* = 1: dark eclipse, grey or brownish in coloration. Surface detail is difficult to distinguish.
- *L* = 2: deep-red or rust-coloured eclipse. Very dark central shadow, whereas outer umbra is relatively bright.
- *L* = 3: brick-red eclipse. Umbral shadow usually has a bright or yellow rim.
- *L* = 4: very bright copper-red or orange eclipse. Umbral shadow has a bluish, very bright rim.

This scale was designed specifically to estimate the brightness of the Moon with the naked eye, which is well suited for our purpose because no high-resolution, technical aids existed in the twelfth and thirteenth centuries. All descriptions of lunar eclipses retrieved and assessed as credible (step 1.1) were rated using the Danjon scale (step 1.2). Note that accounts referring to penumbral and partial eclipses were excluded from analysis as only total lunar eclipse observations are suited to this method[16,17]. The most common adjectives describing lunar eclipses in medieval texts are 'rubeus-a-um' and 'sanguineus-a-um', meaning 'red' and 'blood-coloured', respectively; lunar eclipses so described were rated *L* = 3. A Danjon scale value *L* = 4 was attributed only if the eclipsed Moon was described as exhibiting intense and various colours, such as in this example by the English monk Bartholomew de Cotton of an eclipse on 22 October 1287 ce: "Eodem anno luno in plenilunio visa est crocei, rubei ac varii colori" ("The same year, during the full Moon, the Moon exhibited yellow, red and many other colours"). A Danjon value *L* = 0 was only attributed if the author specifically noted that the Moon had become invisible or extremely dark during the eclipse.

Converting Danjon-scale luminosity estimates to SAOD

To estimate the quantity of volcanic aerosols in the stratosphere and the related attenuation of incoming solar radiation, we converted Danjon's *L* values derived for each lunar eclipse into SAOD values following existing conversion approaches by refs. [16,17,18,19] (step 1.3). The analysis of 46 lunar eclipses that occurred between 1880–1888 and 1960–2001 ce showed that

Danjon values of $L = 4$, 3, 2 and 1 can be closely associated to SAOD values of 0, 0.01, 0.02 and 0.04, respectively, and that an SAOD value of 0.1 or greater is needed for a dark total lunar eclipse ($L = 0$) to occur.

Constraining eruption dates using global aerosol simulations and satellite observations

To determine the period (T_{dark}) when SAOD exceeded 0.1, that is, conditions for a dark total lunar eclipse (step 2.1), we used four (five for Samalas) SAOD time series. For UE1–UE6, we extracted SAOD time series around the 1883 ce Krakatau and 1991 ce Pinatubo eruptions from the Sato/GISS dataset[37]. This dataset, based on satellite observations, ground-based optical measurements and volcanological evidence, reports SAOD at 550 nm since 1850 ce. We also extracted SAOD time series for the Pinatubo eruption from the Global Space-based Stratospheric Aerosol Climatology (GloSSAC v2) dataset[38], which spans the period 1979–2018 ce. As observational data are unavailable before the mid-nineteenth century, we estimated the residence time of volcanic stratospheric aerosols for each eruption (UE1–UE6) from the eVolv2k database[4]. In the case of the 1257 ce Samalas eruptions, we also relied on results of the IPSL climate model (IPSL-CM5A-LR)[40], as it treats aerosol microphysics and has been validated for the well-observed case of the 1991 ce Pinatubo eruption[58].

Each SAOD time series covers a 56-month time window (4 months before the eruption and 52 months after). For eruptions releasing more sulfur than Pinatubo, substantial uncertainties remain about the residence time of aerosols in the stratosphere[44], and the time window for SAOD ≥ 0.1 differs depending on the aerosol model selected (see Methods). Accordingly, for each month within the 56-month time window, we determined the probability for the SAOD = 0.1 threshold to be exceeded on the basis of agreement between time series. Probability was defined as ‘very high’ if all the time series (four for UE1–UE6, five for Samalas) indicated exceedance of the threshold in a given month. Likewise, probability was considered ‘high’ or ‘medium’ if at least three or two, respectively, datasets agreed and as ‘low’ if only one dataset indicated threshold exceedance (step 2.2). SAOD time series provide, for each eruption, the most probable time interval [Month min: Month max] during which a dark lunar eclipse can be

observed after an eruption. Reciprocally, we can assume that the most probable eruption date falls within the time interval [Month max: Month min] before the date of the dark lunar eclipse (step 2.3).

Constraining eruption dates using large-scale tree-ring reconstructions

Abrupt cooling detected in large-scale tree-ring temperature reconstructions has provided independent corroboration of the dating of climatically important volcanic eruptions[59,60]. Here we used the NVOLC v2 (ref. [7]), Sch2015 (ref. [41]) and N-TREND2015 (ref. [42]) reconstructions to refine the dating of HMP eruptions (step 3.1). Following the largest eruptions for which event dates are known (that is, Huaynaputina in February 1600 ce, Parker in January 1641 ce, Tambora in April 1815 ce, Krakatau in August 1883 ce and Pinatubo in June 1991 ce), trees recorded a cooling in Northern Hemisphere summer (June–August, or JJA) temperatures in 1601, 1641, 1816, 1884 and 1992 ce, that is, starting at 17, 6, 14, 10 and 12 months after each eruption, respectively. Accordingly, we consider that cooling observed less than 3 and more than 24 months after an eruption cannot be confidently attributed to volcanic forcing. In agreement with the existing literature[60,61], we assume that peak cooling associated with a volcanic eruption occurs between 6 and 18 months after the eruption, with a highest probability between 9 and 15 months (step 3.2).

Combining evidence for best estimates of eruption timing

In the last steps of our procedure, we estimated the most probable eruption time windows by combining the results from observed and simulated SAOD time series, lunar eclipse (step 2.3) and tree-ring records (step 3.2) using a decision matrix developed as step 4.1 and presented in Extended Data Fig. 1. On the basis of this matrix, we thus considered that an eruption most probably occurred during time windows for which SAOD time series, lunar eclipse and tree-ring records indicate consistently high probabilities (step 4.2).

Testing the four-step procedure using the emblematic 1815 ce Tambora eruption

To test the robustness of our dating approach, we use the well-dated Tambora eruption that occurred on 5 April 1815 ce (Extended Data Fig. 1). Several contemporary sources reported the occurrence of a total lunar eclipse on 9–10 June 1816 ce, among which are the observations made by Capel Lofft (1751–1824 ce) in Ipswich (England) and published in the *Monthly Magazine* in 1816 ce (step 1.1). The darkness of this eclipse impressed contemporary astronomers and was rated $L = 0$ on the Danjon scale[17,62] (step 1.2). Using the conversion scale proposed by refs. [16,17,18,19] (step 1.3), we assume that SAOD had exceeded 0.1 on 9–10 June 1816 ce, that is, 14 months after the Tambora eruption. SAOD observations[37,38] and simulations[39,40] show that the highest probability of SAOD to exceed 0.1 is between 3 and 20 months after an eruption. Therefore, we assume that the most probable eruption time window falls within 3–20 months before the dark lunar eclipse of 9–10 June 1816, that is, between December 1814 and March 1816 ce (step 2.3). Then, we use the abrupt cooling recorded in Northern Hemisphere tree-ring-based temperature reconstructions to refine the time window determined in step 2.3. Peak cooling is generally detected in tree-ring records between 9 to 15 months after a large volcanic eruption. The strong cooling observed in summer 1816 ce in Northern Hemisphere tree-ring reconstructions[7,41,42] therefore points to an eruption occurring between April and October 1815 ce (step 3.2). Finally, we combine probabilities of occurrence estimated from steps 2.3 and 3.2 using a decision matrix (step 4.1) and estimate that the Tambora eruption most probably occurred between May 1815 ce and August 1815 ce (step 4.2). Our estimate thus aligns closely with the actual date of the Tambora eruption (April 1815 ce) and confirms the robustness of our approach, as well as its applicability in the case of HMP eruptions.

Four-step-procedure: uncertainties, caveats and scope for improvement

As with any other methods, the procedure presented in this study to constrain the timing of HMP eruptions comes with several limitations. We address these caveats in the following sections but also present several avenues of research to further refine our estimates.

Challenges in dating volcanic eruptions using ancient lunar eclipse records

Historical observations of lunar eclipse brightness are recognized as a valuable proxy for SAOD following large volcanic eruptions[7,16,17,18,19]. However, precautions are necessary to use this proxy appropriately:

1. i.

   Only total lunar eclipses are suitable. Partial and penumbral eclipses cannot be used for reliable estimates of stratospheric turbidity[16].

2. ii.

   Totality should preferably have been observed in good weather conditions (that is, clear, dark sky), not too near the horizon and not too close to dawn or dusk[16].

3. iii.

   The physical appearance of the Moon during totality must be explicitly described and the colour of the eclipsed disk indicated.

4. iv.

   Reports should be contemporary with the event and preferably by an eyewitness. These conditions are not always met for the medieval sources available (see Supplementary Dataset S1 for more information).

5. v.

   Our study suggests that only lunar eclipses occurring within about 20 months of an eruption are useful for dating purposes and for discriminating the tropospheric versus stratospheric transport of volcanic aerosols. At a given location, the interval between two successive total lunar eclipses ranges between 6 months and 3–4 years (refs. [17,63]). The irregular occurrence of total lunar eclipses can

therefore prevent the dating of a volcanic eruption if the eclipse occurs outside this 20-month period.

6. vi.

   Careful treatment and interpretation is required when studying historical reports of lunar eclipses[64,65], as some descriptions may be too brief or cryptic to provide useful information on stratospheric turbidity and potentially lead to erroneous interpretations. One such example is the lunar eclipse of November 1258 ce recorded in the *Azuma Kagami* (吾妻鏡, vol. 5, pp. 625)[66]:

Shōka 2.10.16

“Clear in the morning. After the hour of the Snake [9 am–11 am], heavy rain and flooding. Houses were swept away and people drowned. At the hour of the Horse [11 am–1 pm] the weather began to clear. During the hour of the Rat [11 pm–1 am] the Moon was eclipsed; it was not properly visible” (see Supplementary Dataset S1).

The description of this event is brief and ambiguous, making it difficult to confidently classify it as a dark lunar eclipse. This account was written several decades after the event and is based on an earlier source that is now lost. As a consequence, we did not attribute any luminosity value to this account.

Despite these challenges, lunar eclipses so far represent the only proxy providing a direct and precise estimate of past atmospheric perturbation by volcanic aerosols. By contrast, sun dimming—references to which have been repeatedly used to identify volcanic dust veils[2,67,68,69,70,71,72,73,74]—can only rarely be dated with comparable accuracy and can also be mistakenly identified (when originating from solar eclipses or solar haloes[75,76]).

Uncertainties in simulating the time evolution of stratospheric aerosols in climate models

An important step in our study is the estimation of a timespan after an eruption in which we expect the stratospheric aerosol to be sufficiently optically thick to cause dark lunar eclipses. This interval, T_{dark}, is computed as the interval when the SAOD exceeds 0.1. We produce a probabilistic estimate of T_{dark} from a combination of observed and simulated global mean SAOD time series. The use of models is necessary because some eruptions included in the study generated much greater SAOD than the largest eruptions of the modern period for which good observations are available. However, this introduces large uncertainties in the time evolution of stratospheric aerosol for the largest sulfur yields. After an initial growth period, SAOD from recent eruptions decays approximately exponentially with time[1]. If this behaviour holds for larger eruptions, then the period during which dark eclipses may occur would lengthen for larger eruptions. On the other hand, models including microphysical processes suggest that larger eruptions produce larger sulfate aerosol particles with consequently shorter stratospheric residence[77,78,79]. If so, it would imply shorter T_{dark} periods. This complexity is reflected in the wide spread in simulated SAOD by an ensemble of state-of-the-art aerosol models in coordinated simulations of the 1815 Tambora eruption[80].

Our analysis considers this uncertainty in SAOD evolution for large sulfur yields. The eVolv2k SAOD time series is produced with the EVA model, which is based on observations of the 1991 Pinatubo eruption and uses only a simple variation of the SAOD decay timescale with eruption magnitude. Comparison of eVolv2k SAOD with the comprehensive aerosol models of the Tambora simulations (Fig. 3 in ref. [80]) shows that the T_{dark} timespan from eVolv2k is comparable with that obtained from models producing the most enduring aerosol perturbations. The large SAOD and aerosol lifetime model differences reflect uncertainties at present about processes for aerosol formation and transport among current state-of-the-art models[44,81]. The eVolv2k SAOD time series depends on the estimated VSSI of each eruption; accordingly, the eVolv2k SAOD time series represents an upper limit for the T_{dark} interval. By contrast, the IPSL model, based on free-running aerosol microphysical processes[82], is seen in the Tambora ensemble to produce one of the fastest SAOD decays and therefore shortest T_{dark} intervals. This behaviour reflects rapid growth of stratospheric sulfate aerosols and greatly

enhanced gravitational settling. The IPSL results therefore provide an estimate of the lower limit on T_{dark}. By including these approximate upper and lower limits on T_{dark} in our analysis, we incorporate the uncertainty in stratospheric aerosol evolution and propagate it into our final constraints on eruption timing.

Caveats on the use of tree-ring-based proxies

Tree rings have been used for nearly 40 years to evaluate the timing and assess the environmental consequences of volcanic eruptions[48,83,84,85,86,87,88,89,90,91,92]. Yet, detecting volcanic signals in tree-ring records comes with several challenges. Several tree-ring parameters have been used to study past volcanic events. One of these is ring width (RW), the annual increments of the growth rings in wood. Trees typically respond to volcanically induced cooling by producing narrow RW. However, the use of this parameter to date and quantify cooling induced by large volcanic eruptions is debated. RW is known to be strongly influenced by biological persistence, which can lead RW temperature-based reconstructions to underestimate, lag and exaggerate the duration of post-eruption cooling[83,90,93,94]. The tree-ring community has therefore advocated for the use of another parameter called 'maximum latewood density' (MXD), regarded as the "gold standard of high-resolution paleoclimatology for temperature reconstructions"[95]. MXD, obtained from high-resolution density profiles measured by X-ray radiodensitometry, is indeed less prone to biological memory and responds more rapidly to climate extremes. However, unfortunately, rather few MXD chronologies extend to before 1300 ce[83,94]. Moreover, volcanic eruptions do not result in globally uniform summer cooling[9,83,90]. Depending on the sulfur yield of the eruption, its latitude and season, prevailing climate conditions and internal variability, some regions will cool, whereas others experience little change[83]. When tree-ring chronologies from various regions are averaged to produce hemispheric temperature reconstructions, the volcanic signal can thus become muted and harder to detect[83].

We used the NVOLC v2 (ref. [7]), Sch2015 (ref. [41]) and N-TREND2015 (ref. [42]) reconstructions to refine the dating of HMP eruptions. The NVOLC v2

reconstruction is composed of 25 tree-ring chronologies (12 MXD and 13 RW chronologies), Sch2015 relies on 15 MXD chronologies distributed across the Northern Hemisphere extratropics, whereas N-TREND2015 is based on a network of 54 records (11 RW, 18 MXD and 25 mixed series composed of RW, MXD and blue intensity records). These reconstructions were selected because they integrate a substantial number of MXD records. We did not use recently published reconstructions that rely exclusively on RW records[96].

Overall comparison of the three Northern Hemisphere reconstructions shows good agreement for UE1–UE6 (Fig. 3 and Extended Data Fig. 2). For UE2, we note that the maximum peak cooling is observed in 1171 ce in N-TREND2015 and one year later in Sch2015 and NVOLC v2. For the 1257 ce Samalas eruption, Sch2015 and N-TREND2015 show less pronounced cooling, which—however—remains in the range of uncertainties of the NVOLC v2 reconstruction. We recognize three sources for these discrepancies. (1) Differences in the tree-ring networks used in the different studies. NVOLC v2 includes only chronologies that encompass the full period between today and the twelfth century, whereas Sch2015 and N-TREND2015 include shorter series as well. (2) Differences in the transfer function used. NVOLC v2 is based on a nested principal component regression—to gradually adjust to a changing number of available proxy records[48,97]—combined with a 1,000-iteration bootstrap approach enabling calculation of uncertainties associated with the reconstruction. By contrast, Sch2015 and N-TREND2015 are based on a scaling approach. (3) The climatological datasets used for calibration. NVOLC v2 uses monthly mean (1805–1972 ce) JJA temperature anomalies (40–90° N) from the recently released Berkeley Earth Surface Temperature (BEST) dataset[98]. Schneider et al.[41] calibrated their proxy records against monthly mean JJA temperature anomalies (1901–1976 ce) derived from the 5° × 5° CRUTEM4v network[99] (30–90°). Wilson et al.[42] scaled their proxy record to CRUTEM4v (40–75° N) MJJA land temperatures during the 1880–1988 ce period. The combination of these differences in tree-ring networks, transfer functions, climatological reference datasets, calibration periods as well as target season inevitably results in differences in the cooling magnitudes for specific events.

Our study relies on state-of-the art reconstructions that efficiently capture post-volcanic summer cooling but several avenues may improve the peak cooling detection and refine the timing of HMP eruptions:

1. i.

   Improvement of the spatial coverage of the millennium-long MXD network with new data from poorly represented regions of the globe.

2. ii.

   Quantitative wood anatomy (QWA). Edwards et al.[100,101] attempted to narrow the period of peak cooling associated with the Laki eruption to late summer 1783 ce using cellular-scale tree-ring proxy measurements. These findings contrast with MXD reconstructions that suggest that the entire 1783 ce summer was exceptionally cold and with tree RW reconstructions that mute the cooling. Such results indicate that QWA data can identify more precisely than tree RW and MXD records the timing of peak cooling following volcanic eruptions within the growing season. Including QWA analyses for the HMP eruptions in the four-step procedure proposed in this study may further refine estimates of eruption timing. Despite promising results, highly resolved QWA is at its early stage. Further, because highly resolved wood anatomical parameter chronologies are costly and labour-intensive, it is unlikely that an operational (Northern Hemisphere) network of QWA records will be available soon.

Discriminating between tropospheric and stratospheric aerosol layers

Most of the refracted sunlight that illuminates the eclipsed Moon passes between 5 and 25 km above the surface of the Earth[19]. Upper tropospheric aerosols (5–10 km) may affect the brightness of the Moon[18] but their residence time is on the order of a few weeks[1,18]. Dark lunar eclipses thus more probably indicate high turbidity of the stratosphere after large volcanic events[16,17,18]. We thus assume that lunar eclipses of reddish or coppery colour (that is, with an L value >1) observed in the aftermath of HMP eruptions indicate that aerosol veils were mainly confined to the troposphere and probably had limited climatic impacts. The robustness of our approach was assessed by comparing our results with sulfur isotope records ($\Delta^{33}S$) from Dome C (Antarctica)[3], which have proven a valuable proxy to distinguish between eruptions whose plumes reached the stratosphere at or above the ozone layer and those that remained below[3,102,103,104,105,106,107].

Data availability

The historical data underlying this study can be found in Supplementary Dataset S1 and is available on Zenodo at https://doi.org/10.5281/zenodo.6907654. The tree-ring-based reconstructions can be downloaded at https://doi.org/10.5281/zenodo.3724674. Ice-core data can be retrieved from the following link and repository: https://doi.org/10.1038/nature14565 and https://doi.org/10.1594/WDCC/eVolv2k_v2. SAOD time series can be obtained from the following links: https://data.giss.nasa.gov/modelforce/strataer/ and https://doi.org/10.1594/WDCC/eVolv2k_v2. Source data are provided with this paper.

Code availability

The codes used in the data processing is available on Zenodo at https://doi.org/10.5281/zenodo.6907654.

References

1. Robock, A. Volcanic eruptions and climate. *Rev. Geophys.* **38**, 191–219 (2000).

2. Sigl, M. et al. Timing and climate forcing of volcanic eruptions for the past 2,500 years. *Nature* **523**, 543–549 (2015).

3. Gautier, E. et al. 2600-years of stratospheric volcanism through sulfate isotopes. *Nat. Commun.* **10**, 466 (2019).

4. Toohey, M. & Sigl, M. Volcanic stratospheric sulfur injections and aerosol optical depth from 500 bce to 1900 ce. *Earth Syst. Sci. Data* **9**, 809–831 (2017).

5. Miller, M. H. et al. Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks. *Geophys. Res. Lett.* **39**, L02708 (2012).

6. Oppenheimer, C. *Eruptions that Shook the World* (Cambridge Univ. Press, 2011).

7. Guillet, S., Corona, C., Ludlow, F., Oppenheimer, C. & Stoffel, M. Climatic and societal impacts of a "forgotten" cluster of volcanic eruptions in 1108-1110 ce. *Sci. Rep.* **10**, 6715 (2020).

8. Lavigne, F. et al. Source of the great A.D. 1257 mystery eruption unveiled, Samalas volcano, Rinjani Volcanic Complex, Indonesia. *Proc. Natl Acad. Sci.* **110**, 16742–16747 (2013).

9. Guillet, S. et al. Climate response to the Samalas volcanic eruption in 1257 revealed by proxy records. *Nat. Geosci.* **10**, 123–128 (2017).

10. Gao, C., Oman, L., Robock, A. & Stenchikov, G. L. Atmospheric volcanic loading derived from bipolar ice cores: accounting for the

spatial distribution of volcanic deposition. *J. Geophys. Res.* **112**, D09109 (2007).

11. Torbenson, M. C. A., Plunkett, G., Brown, D. M., Pilcher, J. R. & Leuschner, H. H. Asynchrony in key Holocene chronologies: evidence from Irish bog pines. *Geology* **43**, 799–802 (2015).

12. Plunkett, G. et al. No evidence for tephra in Greenland from the historic eruption of Vesuvius in 79 ce: implications for geochronology and paleoclimatology. *Clim. Past* **18**, 45–65 (2022).

13. Sigl, M., Toohey, M., McConnell, J. R., Cole-Dai, J. & Severi, M. Volcanic stratospheric sulfur injections and aerosol optical depth during the Holocene (past 11 500 years) from a bipolar ice-core array. *Earth Syst. Sci. Data* **14**, 3167–3196 (2022).

14. Burke, A. et al. Stratospheric eruptions from tropical and extra-tropical volcanoes constrained using high-resolution sulfur isotopes in ice cores. *Earth Planet. Sci. Lett.* **521**, 113–119 (2019).

15. Stothers, R. B. & Rampino, M. R. Historic volcanism, European dry fogs, and Greenland acid precipitation, 1500 B.C. to A.D. 1500. *Science* **222**, 411–413 (1983).

16. Stothers, R. B. Stratospheric transparency derived from total lunar eclipse colors, 1665–1800. *Publ. Astron. Soc. Pac.* **116**, 886–893 (2004).

17. Stothers, R. B. Stratospheric transparency derived from total lunar eclipse colors, 1801–1881. *Publ. Astron. Soc. Pac.* **117**, 1445–1450 (2005).

18. Keen, R. A. Volcanic aerosols and lunar eclipses. *Science* **222**, 1011–1013 (1983).

19. Hofmann, D. et al. in *Geophysical Monograph Series* Vol. 139 (eds. Robock, A. & Oppenheimer, C.) 57–73 (American Geophysical Union, 2003).

20. Stephenson, F. R. *Historical Eclipses And Earth's Rotation* (Cambridge Univ. Press, 1997).

21. Espenak, F. & Meeus, J. *Five Millennium Catalog of Lunar Eclipses: −1999 to +3000* (National Aeronautics and Space Administration, 2009).

22. Liu, Y. T. Eight millennia of eclipses (4000 bce – 4000 ce). http://ytliu.epizy.com/eclipse (2021).

23. Roy, A. E. & Stones, E. L. G. The record of eclipses in the Bury chronicle. *Hist. Res.* **43**, 125–133 (1970).

24. McCluskey, S. C. Gregory of Tours, monastic timekeeping, and early Christian attitudes to astronomy. *Isis* **81**, 8–22 (1990).

25. McCluskey, S. C. *Astronomies and Cultures in Early Medieval Europe* (Cambridge Univ. Press, 1997).

26. Daston, L. & Lunbeck, E. *Histories of Scientific Observation* (Univ. Chicago Press, 2011).

27. Nothaft, C. P. E. *Scandalous Error: Calendar Reform and Calendrical Astronomy in Medieval Europe* (Oxford Univ. Press, 2018).

28. Steele, J. M. *Observations and Predictions of Eclipse Times by Early Astronomers* (Springer, 2000).

29. Stephenson, F. R. in *History of Oriental Astronomy* Vol. 275 (ed. Ansari, S. M. R.) 237–243 (Springer, 2002).

30. McCarthy, D. & Breen, A. Astronomical observations in the Irish annals and their motivation. *Peritia* **11**, 1–43 (1997).

31. Bartlett, R. *The Natural and the Supernatural in the Middle Ages: The Wiles Lecture Given at the Queen's University of Belfast, 2006* (Cambridge Univ. Press, 2008).

32. Smoller, L. A. in *Last Things: Death and the Apocalypse in the Middle Ages* (eds. Bynum, C. W. & Freedman, P. H.) 156–187 (Univ. Pennsylvania Press, 2000).

33. Nothaft, C. P. E. *Walcher of Malvern, De lunationibus and De Dracone* (Brepols, 2017).

34. Pedersen, O. In quest of Sacrobosco. *J. Hist. Astron.* **16**, 175–221 (1985).

35. Danjon, A. Relation entre l'eclairement de la lune eclipsée et l'activité solaire. *L'Astronomie* **35**, 261–265 (1921).

36. Link, F. *Eclipse Phenomena in Astronomy* (Springer, 1969).

37. Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98**, 22987 (1993).

38. Kovilakam, M. et al. The Global Space-based Stratospheric Aerosol Climatology (version 2.0): 1979–2018. *Earth Syst. Sci. Data* **12**, 2607–2634 (2020).

39. Toohey, M., Stevens, B., Schmidt, H. & Timmreck, C. Easy Volcanic Aerosol (EVA v1.0): an idealized forcing generator for climate simulations. *Geosci. Model Dev.* **9**, 4049–4070 (2016).

40. Dufresne, J.-L. et al. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Clim. Dyn.* **40**, 2123–2165 (2013).

41. Schneider, L. et al. Revising midlatitude summer temperatures back to A.D. 600 based on a wood density network: revising hemispheric temperature history. *Geophys. Res. Lett.* **42**, 4556–4562 (2015).

42. Wilson, R. et al. Last millennium northern hemisphere summer temperatures from tree rings: part I: the long term context. *Quat. Sci. Rev.* **134**, 1–18 (2016).

43. Bauch, M. in *The Dance of Death in Late Medieval and Renaissance Europe: Environmental Stress, Mortality and Social Response* (eds. Kiss, A. & Pribyl, K.) 241 (Routledge, Taylor & Francis, 2020).

44. Clyne, M. et al. Model physics and chemistry causing intermodel disagreement within the VolMIP-Tambora Interactive Stratospheric Aerosol ensemble. *Atmos. Chem. Phys.* **21**, 3317–3343 (2021).

45. Stevenson, S., Fasullo, J. T., Otto-Bliesner, B. L., Tomas, R. A. & Gao, C. Role of eruption season in reconciling model and proxy responses to tropical volcanism. *Proc. Natl Acad. Sci.* **114**, 1822–1826 (2017).

46. Toohey, M., Krüger, K., Niemeier, U. & Timmreck, C. The influence of eruption season on the global aerosol evolution and radiative impact of tropical volcanic eruptions. *Atmos. Chem. Phys.* **11**, 12351–12367 (2011).

47. Otto-Bliesner, B. L. et al. Climate variability and change since 850 ce: an ensemble approach with the Community Earth System Model. *Bull. Am. Meteorol. Soc.* **97**, 735–754 (2016).

48. Stoffel, M. et al. Estimates of volcanic-induced cooling in the Northern Hemisphere over the past 1,500 years. *Nat. Geosci.* **8**, 784–788 (2015).

49. Sorimachi, S. *Kōbunsō Keiaisho Zuroku. II* (*弘文荘敬愛書図録. II, Images of Respected and Beloved Books of Kōbunsō, Vol. 2*) (Kōbunsō, 1984).

50. Atkins, P. S. Meigetsuki, the diary of Fujiwara no Teika: Karoku 2.9 (1226). *J. Am. Orient. Soc.* **130**, 235–258 (2010).

51. Vyssotsky, A. N. *Astronomical Records in the Russian Chronicles from 1000 to 1600 A. D.* (*as Collected by D. O. Sviatsky*) (Lund Observatory, 1949).

52. Stephenson, F. R. & Said, S. Records of lunar eclipses in medieval Arabic chronicles. *Bull. Sch. Orient. Afr. Stud.* **60**, 1–34 (1997).

53. Kanda, S. *Nihon Tenmon Shiryô* (*日本天文史料, Japanese Historical Records of Astronomy*) (1935).

54. Stephenson, F. R., Morrison, L. V. & Hohenkerk, C. Y. Measurement of the Earth's rotation: 720 BC to AD 2015. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **472**, 20160404 (2016).

55. Morrison, L. V., Stephenson, F. R., Hohenkerk, C. Y. & Zawilski, M. Addendum 2020 to 'Measurement of the Earth's rotation: 720 BC to AD 2015'. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **477**, 20200776 (2021).

56. Sawyer Hogg, H. Out of Old Books (Lunar Eclipses). *J. R. Astron. Soc. Can.* **40**, 230–232 (1946).

57. Stothers, R. B. Cloudy and clear stratospheres before A.D. 1000 inferred from written sources. *J. Geophys. Res. Atmos.* **107**, AAC-17 (2002).

58. Bekki, S. & Pyle, J. A. Two-dimensional assessment of the impact of aircraft sulphur emissions on the stratospheric sulphate aerosol layer. *J. Geophys. Res. Atmos.* **97**, 15839–15847 (1992).

59. Baillie, M. G. L. & McAneney, J. Tree ring effects and ice core acidities clarify the volcanic record of the first millennium. *Clim. Past* **11**, 105–114 (2015).

60. Schneider, L., Smerdon, J. E., Pretis, F., Hartl-Meier, C. & Esper, J. A new archive of large volcanic events over the past millennium derived from reconstructed summer temperatures. *Environ. Res. Lett.* **12**, 094005 (2017).

61. Tejedor, E., Steiger, N. J., Smerdon, J. E., Serrano-Notivoli, R. & Vuille, M. Global hydroclimatic response to tropical volcanic eruptions over the last millennium. *Proc. Natl Acad. Sci. USA* **118**, e2019145118 (2021).

62. Stothers, R. B. The great Tambora eruption in 1815 and its aftermath. *Science* **224**, 1191–1198 (1984).

63. Stothers, R. B. Three centuries of observation of stratospheric transparency. *Clim. Change* **83**, 515–521 (2007).

64. Bell, W. T. & Ogilvie, A. E. J. Weather compilations as a source of data for the reconstruction of European climate during the medieval period. *Clim. Change* **1**, 331–348 (1978).

65. Ogilvie, A. E. J. & Farmer, G. Documenting the medieval climate. in *Climates of the British Isles: Present, Past, and Future* (eds Hulme, M. & Barrow, E.) 112–133 (Routledge, 1997).

66. *Azuma Kagami* (*吾妻鏡*). Kokushi taikei 国史大系, Vol. 5 (1903).

67. Lamb, H. H. Volcanic dust in the atmosphere; with a chronology and assessment of its meteorological significance. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **266**, 425–533 (1970).

68. Stothers, R. B. Mystery cloud of AD 536. *Nature* **307**, 344–345 (1984).

69. Stothers, R. B. The great dry fog of 1783. *Clim. Change* **32**, 79–89 (1996).

70. Stothers, R. B. Volcanic dry fogs, climate cooling, and plague pandemics in Europe and the Middle East. *Clim. Change* **42**, 713–723 (1999).

71. Grattan, J. P. & Pyatt, F. B. Volcanic eruptions dry fogs and the European palaeoenvironmental record: localised phenomena or hemispheric impacts? *Global Planet. Change* **21**, 173–179 (1999).

72. Guevara-Murua, A., Williams, C. A., Hendy, E. J., Rust, A. C. & Cashman, K. V. Observations of a stratospheric aerosol *veil* from a tropical volcanic eruption in December 1808: is this the *Unknown* ~1809 eruption? *Clim. Past* **10**, 1707–1722 (2014).

73. Oppenheimer, C. et al. The Eldgjá eruption: timing, long-range impacts and influence on the Christianisation of Iceland. *Clim. Change* **147**, 369–381 (2018).

74. McConnell, J. R. et al. Extreme climate after massive eruption of Alaska's Okmok volcano in 43 bce and effects on the late Roman Republic and Ptolemaic Kingdom. *Proc. Natl Acad. Sci. USA* **117**, 15443–15449 (2020).

75. Brugnatelli, V. & Tibaldi, A. Effects in North Africa of the 934–940 ce Eldgjá and 1783–1784 ce Laki eruptions (Iceland) revealed by previously unrecognized written sources. *Bull. Volcanol.* **82**, 73 (2020).

76. Newfield, T. P. & Oppenheimer, C. Comment on "Effects in North Africa of the 934–940 ce Eldgjá and 1783–1784 ce Laki eruptions (Iceland) revealed by previously unrecognized written sources" by Brugnatelli, V., and Tibaldi, A. [Bull. Volcanol. (2020) 82:73]. *Bull. Volcanol.* **83**, 55 (2021).

77. Pinto, J. P., Turco, R. P. & Toon, O. B. Self-limiting physical and chemical effects in volcanic eruption clouds. *J. Geophys. Res.* **94**, 11165 (1989).

78. Timmreck, C. et al. Limited temperature response to the very large AD 1258 volcanic eruption. *Geophys. Res. Lett.* **36**, L21708 (2009).

79. Timmreck, C. et al. Aerosol size confines climate response to volcanic super-eruptions. *Geophys. Res. Lett.* **37**, L24705 (2010).

80. Zanchettin, D. et al. The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geosci. Model Dev.* **9**, 2701–2719 (2016).

81. Marshall, L. et al. Multi-model comparison of the volcanic sulfate deposition from the 1815 eruption of Mt. Tambora. *Atmos. Chem. Phys.* **18**, 2307–2328 (2018).

82. Bekki, S. Oxidation of volcanic SO_2: a sink for stratospheric OH and H_2O. *Geophys. Res. Lett.* **22**, 913–916 (1995).

83. D'Arrigo, R., Wilson, R. & Anchukaitis, K. J. Volcanic cooling signal in tree ring temperature records for the past millennium. *J. Geophys. Res. Atmos.* **118**, 9000–9010 (2013).

84. LaMarche, V. C. & Hirschboeck, K. K. Frost rings in trees as records of major volcanic eruptions. *Nature* **307**, 121–126 (1984).

85. Scuderi, L. A. Tree-ring evidence for climatically effective volcanic eruptions. *Quat. Res.* **34**, 67–85 (1990).

86. Lough, J. M. & Fritts, H. C. An assessment of the possible effects of volcanic eruptions on North American climate using tree-ring data, 1602 to 1900 A.D. *Clim. Change* **10**, 219–239 (1987).

87. Jones, P. D., Briffa, K. R. & Schweingruber, F. H. Tree-ring evidence of the widespread effects of explosive volcanic eruptions. *Geophys. Res. Lett.* **22**, 1333–1336 (1995).

88. Briffa, K. R., Jones, P. D., Schweingruber, F. H. & Osborn, T. J. Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years. *Nature* **393**, 450–455 (1998).

89. D'Arrigo, R., Jacoby, G. & Frank, D. in *Geophysical Monograph Series* Vol. 139 (eds. Robock, A. & Oppenheimer, C.) 255–261 (American Geophysical Union, 2003).

90. Anchukaitis, K. J. et al. Tree rings and volcanic cooling. *Nat. Geosci.* **5**, 836–837 (2012).

91. Büntgen, U. et al. Cooling and societal change during the Late Antique Little Ice Age from 536 to around 660 AD. *Nat. Geosci.* **9**, 231–236 (2016).

92. Büntgen, U. et al. Prominent role of volcanism in Common Era climate variability and human history. *Dendrochronologia* **64**, 125757

(2020).

93. Frank, D., Büntgen, U., Böhm, R., Maugeri, M. & Esper, J. Warmer early instrumental measurements versus colder reconstructed temperatures: shooting at a moving target. *Quat. Sci. Rev.* **26**, 3298–3310 (2007).

94. Esper, J., Schneider, L., Smerdon, J. E., Schöne, B. R. & Büntgen, U. Signals and memory in tree-ring width and density data. *Dendrochronologia* **35**, 62–70 (2015).

95. Björklund, J. et al. Scientific merits and analytical challenges of tree-ring densitometry. *Rev. Geophys.* **57**, 1224–1264 (2019).

96. Büntgen, U. et al. The influence of decision-making in tree ring-based climate reconstructions. *Nat. Commun.* **12**, 3411 (2021).

97. Ortega, P. et al. A model-tested North Atlantic Oscillation reconstruction for the past millennium. *Nature* **523**, 71–74 (2015).

98. Rohde, R., Muller, R. A., Jacobsen, R., Muller, E. & Wickham, C. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinfor. Geostat. An Overview* **1**, 1 (2013).

99. Jones, P. D. et al. Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010. *J. Geophys. Res. Atmos.* **117**, D05127 (2012).

100. Edwards, J. et al. Intra-annual climate anomalies in northwestern North America following the 1783–1784 ce Laki eruption. *Geophys. Res. Atmos.* **126**, e2020JD033544 (2021).

101. Edwards, J. et al. The origin of tree-ring reconstructed summer cooling in northern Europe during the 18th century eruption of Laki. *Paleoceanogr. Paleoclimatol.* **37**, e2021PA004386 (2022).

102. Savarino, J., Romero, A., Cole-Dai, J., Bekki, S. & Thiemens, M. H. UV induced mass-independent sulfur isotope fractionation in

stratospheric volcanic sulfate. *Geophys. Res. Lett.* **30**, 2131 (2003).

103. Baroni, M., Savarino, J., Cole-Dai, J., Rai, V. K. & Thiemens, M. H. Anomalous sulfur isotope compositions of volcanic sulfate over the last millennium in Antarctic ice cores. *J. Geophys. Res.* **113**, D20112 (2008).

104. Cole-Dai, J. et al. Cold decade (AD 1810–1819) caused by Tambora (1815) and another (1809) stratospheric volcanic eruption. *Geophys. Res. Lett.* **36**, L22703 (2009).

105. Lanciki, A., Cole-Dai, J., Thiemens, M. H. & Savarino, J. Sulfur isotope evidence of little or no stratospheric impact by the 1783 Laki volcanic eruption. **39**, L01806 (2012).

106. Cole-Dai, J. et al. Two likely stratospheric volcanic eruptions in the 1450s C.E. found in a bipolar, subannually dated 800 year ice core record. *J. Geophys. Res. Atmos.* **118**, 7459–7466 (2013).

107. Crick, L. et al. New insights into the ~74 ka Toba eruption from sulfur isotopes of polar ice cores. *Clim. Past* **17**, 2119–2137 (2021).

108. Kepler, J. *Ad vitellionem paralipomena, quibus astronomiae pars optica traditur; potissimùm de artificiosa observatione et aestimatione diametrorum delinquiorumq; solis & lunae. Cum exemplis insignium eclipsium. In tractatum luculentum de modo visionis, & humorum oculi usu, contra opticos & anatomicos* (Apud Claudium Marnium & Haeredes FIoannis Aubrii, 1604).

109. Link, F. in *Physics and Astronomy of the Moon* (ed. Kopal, Z.) 161–229 (Academic, 1961).

110. Hevelius, J. *Johannis Hevelii Selenographia sive lunae descriptio: atque accurata, tam macularum ejus, ... delineatio, ...; addita est, lentes expoliendi nova ratio, ut et telescopia diversa construendi, et experiendi, ... explicatur* (Hünefeld, 1647).

111. Flammarion, C. L'éclipse totale de Lune du 4 octobre. *L'Astronomie* **3**, 401–408 (1884).

112. Flammarion, C. L'éclipse de Lune du 4 octobre 1884. *L'Astronomie* **5**, 23–30 (1885).

113. Sawyer Hogg, H. Out of Old Books (The Lunar Eclipse of October 4, 1884). *J. R. Astron. Soc. Can.* **40**, 267–271 (1946).

114. Flint, A. S. The Total Lunar Eclipse of March 22, 1913. *Pop. Astron.* **22**, 425–427 (1914).

115. De Vaucouleurs, G. La loi normale de luminosité des eclipses de Lune de 1894 à 1943. *C. R. Acad. Sci.* **218**, 655–657 (1944).

116. Volz, F. E. Atmospheric turbidity after the Agung eruption of 1963 and size distribution of the volcanic aerosol. *J. Geophys. Res.* **75**, 5185–5193 (1970).

117. Link, F. La loi de Danjon et les eruptions volcaniques. *Moon* **11**, 261–272 (1974).

118. Haas, W. H. A grey moon dim in the umbra: the total lunar eclipse of December 30, 1982. *Strolling Astron.* **29**, 259–264 (1983).

119. Schaaf, F. *The Starry Room: Naked Eye Astronomy in the Intimate Universe* (Wiley, 1988).

120. O'Meara, S. J. & di Cicco, D. Making sense of November's perplexing lunar eclipse. *Sky Telesc.* **87**, 103–107 (1994).

Acknowledgements

S.G., C.C., M.K. and M. Stoffel were supported by the Swiss National Science Foundation Sinergia Project CALDERA (CRSII5_183571). S.G. acknowledges A. Harrak (Department of Near and Middle Eastern Civilizations, University of Toronto), F. Espenak (NASA Goddard Space Flight Center), F. Hierink (Institute for Environmental Sciences, University

of Geneva) and P. Souyri (Department of East Asian Studies, University of Geneva) for providing advice on the manuscript. F. Lavigne was supported by Institut Universitaire de France (IUF, Academic Institute of France). M.K. received funding from the EUR IPSL – Climate Graduate School project, managed by the ANR within the "Investissements d'avenir" programme under reference ANR-11-IDEX-0004-17-EURE-0006. F. Ludlow received funding from an Irish Research Council Starting Laureate Award (CLICAB project, IRCLA/2017/303). F. Ludlow and Z.Y. also received funding from a European Research Council (ERC) Synergy Grant (4-OCEANS; grant agreement no. 951649) under the European Union's Horizon 2020 research and innovation programme. M. Sigl received funding from the ERC under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 820047). This paper is a product of the Volcanic Impacts on Climate and Society (VICS) working group.

Funding

Open access funding provided by University of Geneva.

Author information

Authors and Affiliations

1. Climate Change Impacts and Risks in the Anthropocene (C-CIA), Institute for Environmental Sciences, University of Geneva, Geneva, Switzerland

   Sébastien Guillet, Christophe Corona, Tomoko Muranaka & Markus Stoffel

2. GEOLAB, Université Clermont Auvergne, CNRS, Clermont-Ferrand, France

   Christophe Corona

3. Department of Geography, University of Cambridge, Cambridge, UK

Clive Oppenheimer

4. Laboratoire de Géographie Physique, Université Paris 1 Panthéon-Sorbonne, Thiais, France

Franck Lavigne

5. Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques, IPSL, Sorbonne Université/IRD/CNRS/MNHN, Paris, France

Myriam Khodri

6. Trinity Centre for Environmental Humanities, Department of History, School of Histories & Humanities, Trinity College Dublin, Dublin, Ireland

Francis Ludlow & Zhen Yang

7. Climate and Environmental Physics, University of Bern, Bern, Switzerland

Michael Sigl

8. Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

Michael Sigl

9. Department of Physics and Engineering Physics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Matthew Toohey

10. Department of Asian Languages & Literature, University of Washington, Seattle, WA, USA

Paul S. Atkins & Nobuko Horikawa

11. Department of Earth Sciences, University of Geneva, Geneva, Switzerland

    Markus Stoffel

12. Department F.-A. Forel for Environmental and Aquatic Sciences, University of Geneva, Geneva, Switzerland

    Markus Stoffel

Contributions

S.G. designed the research, with inputs from C.C., M. Stoffel and F. Lavigne. S.G. investigated European, Russian and Middle Eastern historical sources. S.G., P.S.A., N.H. and T.M. investigated Japanese historical sources. Z.Y. and S.G. investigated Chinese historical sources. S.G. and Z.Y. investigated Korean historical sources. S.G. analysed historical sources, with contributions from Z.Y., P.S.A. and F. Ludlow. S.G., C.C., M. Sigl, C.O. and M.T. contributed to ice-core and tree-ring data interpretation. S.G., C.C., M.K. and M.T. contributed to aerosol model simulations interpretation. The manuscript was written by S.G., C.C., M. Stoffel and C.O., with contributions from M.K., F. Ludlow, F. Lavigne, M. Sigl and M.T.

Corresponding author

Correspondence to Sébastien Guillet.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Harri Kokkola, Anne Mathers-Lawrence and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Testing the four-step procedure using the emblematic 1815 ce Tambora eruption.

A detailed description of the approach can be found in Methods.

Source data

Extended Data Fig. 2 Potential dates of HMP (1100–1300 ce) volcanic eruptions (UE1 and UE5) using a four-step procedure.

a, Residence time of volcanic stratospheric aerosols and time windows with SAOD exceeding about 0.1. The residence time of aerosols is based on global mean $SAOD_{550}$ time series from the Sato/GISS[37] and GloSSAC v2 (ref. [38]) datasets (for the 1883 ce Krakatau and 1991 ce Pinatubo eruptions) and simulated by the EVA forcing generator[4,39] (for UE1 and UE5). Probability of occurrence of HMP eruptions based on the timing of dark lunar eclipse dates (**b**) and tree-ring records[7,41,42] (**c**). **d**, Integration of **b** and **c** to estimate the most probable time windows for the UE1 and UE5 eruptions.

Source data

Extended Data Table 1 Number and percentage of total lunar eclipses potentially visible and recorded in historical sources in Europe during the twelfth and thirteenth centuries

Extended Data Table 2 Description of lunar eclipses observed after the eight largest eruptions since 1600 ce (refs. [16,17,19,62,108,109,110,111,112,113,114,115,116,117,118,119,120])

Supplementary information

Peer Review File

Supplementary Dataset S1

Database of lunar eclipses retrieved from historical sources from Europe, the Middle East and East Asia for the twelfth and thirteenth centuries

Supplementary Dataset S2

Summary table describing the volcanic eruptions considered in this study

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Guillet, S., Corona, C., Oppenheimer, C. *et al.* Lunar eclipses illuminate timing and climate impact of medieval volcanism. *Nature* **616**, 90–95 (2023). https://doi.org/10.1038/s41586-023-05751-z

- Received: 07 October 2021
- Accepted: 20 January 2023
- Published: 05 April 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05751-z

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

The medieval Moon unveils volcanic secrets

- Andrea Seim
- Eduardo Zorita
- Anne Lawrence-Mathers

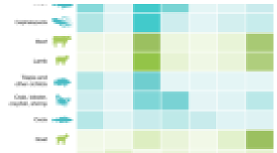
Nature News & Views Forum 05 Apr 2023

Medieval accounts of eclipses shine light on massive volcanic eruptions

- Jeff Tollefson

Nature News 05 Apr 2023

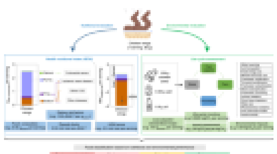
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05751-z



Aquatic foods to nourish nations

15 September 2021

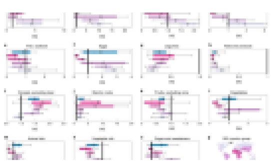
Christopher D. Golden, J. Zachary Koehn, … Shakuntala H. Thilsted



Small targeted dietary changes can yield substantial gains for human health and the environment

18 August 2021

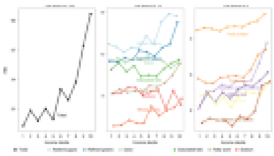
Katerina S. Stylianou, Victor L. Fulgoni III & Olivier Jolliet



Compliance with EAT–Lancet dietary guidelines would reduce global water footprint but increase it for 40% of the world population

31 January 2022

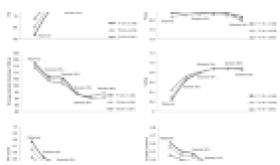
Marta Tuninetti, Luca Ridolfi & Francesco Laio



Shifts towards healthy diets in the US can reduce environmental impacts but would be unaffordable for poorer minorities

30 August 2021

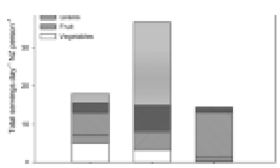
Pan He, Kuishuang Feng, … Klaus Hubacek



Conservative to disruptive diets for optimizing nutrition, environmental impacts and cost in French adults from the NutriNet-Santé cohort

11 March 2021

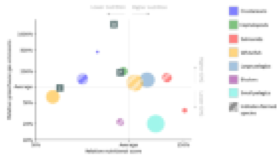
Louise Seconda, Helene Fouillet, … Emmanuelle Kesse-Guyot



Food exports and imports of New Zealand in relation to the food-based dietary guidelines

10 January 2020

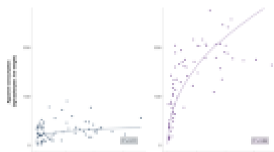
Elaine Rush & Vladimir Obolonkin



[Assessing seafood nutritional diversity together with climate impacts informs more comprehensive dietary advice]

08 September 2022

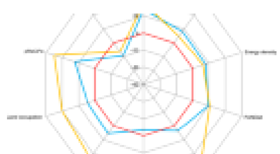
Marta Bianchi, Elinor Hallström, … Friederike Ziegler



[Blue food demand across geographic and temporal scales]

15 September 2021

Rosamond L. Naylor, Avinash Kishore, … Beatrice Crona



[Sustainability analysis of French dietary guidelines using multiple criteria]

23 March 2020

Emmanuelle Kesse-Guyot, Dan Chaltiel, … Serge Hercberg

- Article
- Open Access
- Published: 22 February 2023

Four ways blue foods can help achieve food system ambitions across nations

- Beatrice I. Crona ,
- Emmy Wassénius,
- Malin Jonell,
- J. Zachary Koehn,
- Rebecca Short,
- Michelle Tigchelaar,
- Tim M. Daw,
- Christopher D. Golden,
- Jessica A. Gephart,
- Edward H. Allison,
- Simon R. Bush,
- Ling Cao,
- William W. L. Cheung,
- Fabrice DeClerck,
- Jessica Fanzo,
- Stefan Gelcich,
- Avinash Kishore,
- Benjamin S. Halpern,
- Christina C. Hicks,
- James P. Leape,
- David C. Little,
- Fiorenza Micheli,
- Rosamond L. Naylor,
- Michael Phillips,
- Elizabeth R. Selig,
- Marco Springmann,
- U. Rashid Sumaila,
- Max Troell,
- Shakuntala H. Thilsted &
- …

- Colette C. C. Wabnitz

Show authors

Nature **volume 616**, pages 104–112 (2023)

- 12k Accesses
- 241 Altmetric
- Metrics details

Abstract

Blue foods, sourced in aquatic environments, are important for the economies, livelihoods, nutritional security and cultures of people in many nations. They are often nutrient rich[1], generate lower emissions and impacts on land and water than many terrestrial meats[2], and contribute to the health[3], wellbeing and livelihoods of many rural communities[4]. The Blue Food Assessment recently evaluated nutritional, environmental, economic and justice dimensions of blue foods globally. Here we integrate these findings and translate them into four policy objectives to help realize the contributions that blue foods can make to national food systems around the world: ensuring supplies of critical nutrients, providing healthy alternatives to terrestrial meat, reducing dietary environmental footprints and safeguarding blue food contributions to nutrition, just economies and livelihoods under a changing climate. To account for how context-specific environmental, socio-economic and cultural aspects affect this contribution, we assess the relevance of each policy objective for individual countries, and examine associated co-benefits and trade-offs at national and international scales. We find that in many African and South American nations, facilitating consumption of culturally relevant blue food, especially among nutritionally vulnerable population segments, could address vitamin B_{12} and omega-3 deficiencies. Meanwhile, in many global North nations, cardiovascular disease rates and large greenhouse gas footprints from ruminant meat intake could be lowered through moderate consumption of seafood with low environmental impact. The analytical framework we

provide also identifies countries with high future risk, for whom climate adaptation of blue food systems will be particularly important. Overall the framework helps decision makers to assess the blue food policy objectives most relevant to their geographies, and to compare and contrast the benefits and trade-offs associated with pursuing these objectives.

Main

Given the diverse contribution of blue foods to society, the role that they can play in the transition to healthier, more just and less environmentally harmful food systems is an important question for both public and private decision makers. Yet, blue foods have remained remarkably absent from many contemporary food system discussions and policies on both nature and nutrition-positive outcomes[5,6,7,8]. When included, their representation is often simplified and reduced to a few types of 'fish' in dietary recommendations[9] and demand projections[10]. Similarly, ocean policy often neglects blue food contributions to human nutrition and benefits to communities producing them[11,12]. Deeper appreciation and understanding of the roles blue foods can play is essential for informing policy development that can harness their unique capacity for addressing nutritional, social and environmental food system challenges, while navigating the trade-offs of pursuing these different roles, within and across countries.

Blue foods are immensely diverse. More than 2,200 wild species are caught and more than 600 are farmed[13], with tremendous variation in associated production and processing systems and practices[2,14]. Aquatic food consumption profiles of nations are also remarkably diverse[10]. This diversity means that blue foods vary substantially in their contributions to human health, nutrition[1], jobs[15] and culture[16] and their environmental impacts[2]. Natural variations in blue food diversity and abundance are compounded by social structures that exacerbate inequities[17] across socio-economic and geographic contexts. Diversity can bolster the resilience of blue foods to shocks[18,19], but such resilience is unevenly represented across countries at present[20].

In this diversity lies the key to understanding the geographic contexts and conditions whereby blue foods can contribute to achieving food system ambitions, such as improved nutrition, equity and lowered environmental impact—as articulated by high-level processes[21,22] and the Sustainable Development Goals of the United Nations[23].

This paper integrates the findings of an initiative to assess the multiple roles animal-sourced blue foods play in food systems around the world (the Blue Food Assessment; https://www.bluefood.earth/) and translates them into four policy objectives that could help realize the positive health, environment, resilience and equity contributions of aquatic foods worldwide. We assess the relevance of these policy objectives for individual countries, and then examine the co-benefits and trade-offs associated with policy objectives at national and international scales. In doing so, we provide a guiding framework for decision makers across public and private spheres to assess blue food policy objectives most relevant to their geographies, and compare and contrast the benefits and trade-offs associated with pursuing these objectives.

Four ways blue foods improve food systems

The Blue Food Assessment examined the roles of blue foods in current and future food systems globally. It brought together more than 100 scholars across a wide range of disciplines to investigate the nutritional contribution of blue foods[1], current and future demand[10], and the environmental impacts of blue food production[2], as well as the vulnerability of this production to environmental stressors[24] (L.C., manuscript in preparation). It synthesized key dimensions characterizing the small-scale fisheries and aquaculture (SSFA) actors[14] who produce two-thirds of aquatic foods destined for human consumption[14,25], and evaluated injustices across the blue food system to identify policy attributes that support more equitable access to blue food benefits[17]. It also assessed the climate risks posed to nutritional, social, economic and environmental outcomes of blue food systems worldwide[20]. Finally, it explored how supporting the capabilities of actors, small and large, across the supply chains can build adaptive capacity to support a wider food system transformation (S.R.B., manuscript in

preparation). This multi-perspective assessment is unique, and together with the large body of previous research helps crystallize the diverse functions blue foods play at present, and how these can be leveraged to support a food system transformation. These functions include the following.

Sources of critical nutrients

Blue foods are rich in many essential nutrients[26]. Like other animal-source foods, blue foods can enhance bioavailability of nutrients in plant-based food sources, depending on how they are combined with other foods[27]. Where blue foods are accessible and consumed in adequate quantities, they can promote nutrition by reducing deficiencies of a range of nutrients, most notably vitamin B_{12} and the omega-3 long-chain polyunsaturated fatty acids docosahexaenoic acid and eicosapentaenoic acid (DHA and EPA; hereafter, fatty acids), in which blue foods are generally rich. These are among the nutrients noted as important for human nutrition[21], showing relatively high levels of deficiency globally[1] (Extended Data Fig. 1), and blue foods are projected to contribute a global average of approximately 27% and 100% of omega-3 fatty acids and vitamin B_{12}, respectively, by 2030 (ref. [1]). Addressing these deficiencies is particularly important among vulnerable demographic groups, such as young children and older people, pregnant women and women of childbearing age[28,29]. Alongside other health-critical foods, blue foods can thus make essential contributions to maintaining and improving nutritional food system outcomes[30]. Capture fisheries constitute the last large wild-food resource. Failing to sustain it will jeopardize food security in many places and it will be challenging to replace without negative environmental consequences.

Healthy alternatives to terrestrial animal-source foods

By adding to the range of food sources associated with relative reductions in many non-communicable diseases[31,32,33], blue foods can help to circumvent the harmful nutrition transition observed in many countries at present (sensu ref. [34]), and contribute to reducing the overall disease burden. This may be particularly relevant in countries experiencing continued high, or growing, trends of red (particularly processed) meat intake (such as China, Argentina,

Brazil, the USA and Eastern Europe)[35,36,37]. Cardiovascular disease is among the most commonly cited negative health effects of red meat consumption[32,33], and we use it here as an example of how countries can assess the relevance of blue food policies depending on their specific disease burden. In this context, the health-promoting role of blue foods rests on the assumption that they can displace some red meat consumption[1,10] and on the plausible health contributions (for example, of DHA and EPA from aquatic foods[38,39]), for which uncertainties persist[39,40,41]. Substitutability of red meat by fish has not been well documented, yet reverse substitution has been observed[30], as have large-scale adoptions of new proteins when innovations are supported with public funds, and scaled by the private sector under supportive state and international policy regimes[42]. Sixty years of increased consumption of poultry compared to beef[10] suggests that poultry and seafood can replace red meat. As blue foods are already part of the local food culture in many countries with a high level of meat consumption, they constitute a promising step away from routinized overconsumption of red meat.

Nutrient sources with relatively low environmental footprints

Across the diversity of blue foods, many production systems already result in relatively lower environmental pressures compared to those associated with terrestrial animal-source food production[2]. Partial replacement of particularly ruminant meat with blue foods can therefore help to lower dietary environmental footprints. Unfed aquaculture systems, such as bivalves and seaweeds, typically result in low greenhouse gas (GHG), nitrogen and phosphorus emissions and require limited freshwater and land inputs. Many fed aquaculture systems perform similarly to or better than chicken production, which is often considered the most efficient terrestrial animal-source food production system[2,43]. GHG emissions for capture fisheries vary substantially[44], with small pelagic fish, cod and some inland fisheries resulting in low average emissions[45] and flounder and lobsters having high emissions[2], but all capture fisheries generally have negligible N and P emissions, and freshwater and land inputs[2,46]. Blue food production can nonetheless restructure ecological food webs and cause substantial biodiversity loss[47,48], but there is a large potential to reduce most

environmental impacts associated with blue food production. Improved fisheries management, fossil-free energy and a shift to low-impact gear are key areas of interventions for capture fisheries[49]. Impacts from aquaculture could be substantially reduced by lowering feed conversion ratios (for example, through breeding programmes), shifting species focus or feed composition (for example, to deforestation-free soy, fisheries by-products, or insect meal) and improving husbandry practices[2,50,51].

Cornerstones in cultures, diets, economies and livelihoods

In many nations, blue foods are a cornerstone of cultures, diets and economies, fulfilling critical food and nutrition security functions[4]. Blue foods are also among the most traded commodities globally, providing substantial export revenue for many nations[13] and livelihoods for 800 million people[25], indicating their critical role for employment and subsistence. However, although blue foods support the welfare (for example, through jobs and nutrient-rich blue food) of these actors[17], the wealth-generating benefits (for example, export revenues) of blue foods flow predominantly to industrial-scale firms that control global supply chains[17,52,53]. This reflects inequities inherent across many food systems[54]. Small-scale actors are therefore often undervalued and marginalized in decision making, threatening livelihoods and their capacity to cope with changing environmental conditions[17,55]. Policies focusing on environmental or economic gains must therefore be attentive to risks of inadvertently undermining human wellbeing. Several environmental stressors affect blue food production, and climate change in particular will affect all aspects of aquatic food systems, from production to consumption[20]. Overall, the climate risk to a country's aquatic food system is determined not only by the climate hazards the country faces, but also by its dependence on the nutritional, economic, social and environmental benefits of aquatic foods[56], and the vulnerability to losing these benefits[20]. These future threats may compound existing challenges and exacerbate inequities, by increasing barriers to inclusive production and trade, limiting access to blue foods, and thus restricting their nutritional contributions[14,20,57]. Supporting the diversity and resilience of SSFA[14] can help build national food system resilience to climate and other shocks[14,20,58], by providing response diversity[59].

Anticipation of how and where climate hazards will be most severe is therefore essential to help private and public actors identify appropriate actions to safeguard the contribution of blue food to the health, economies, culture and livelihoods in a way that also considers justice.

From science to policy objectives

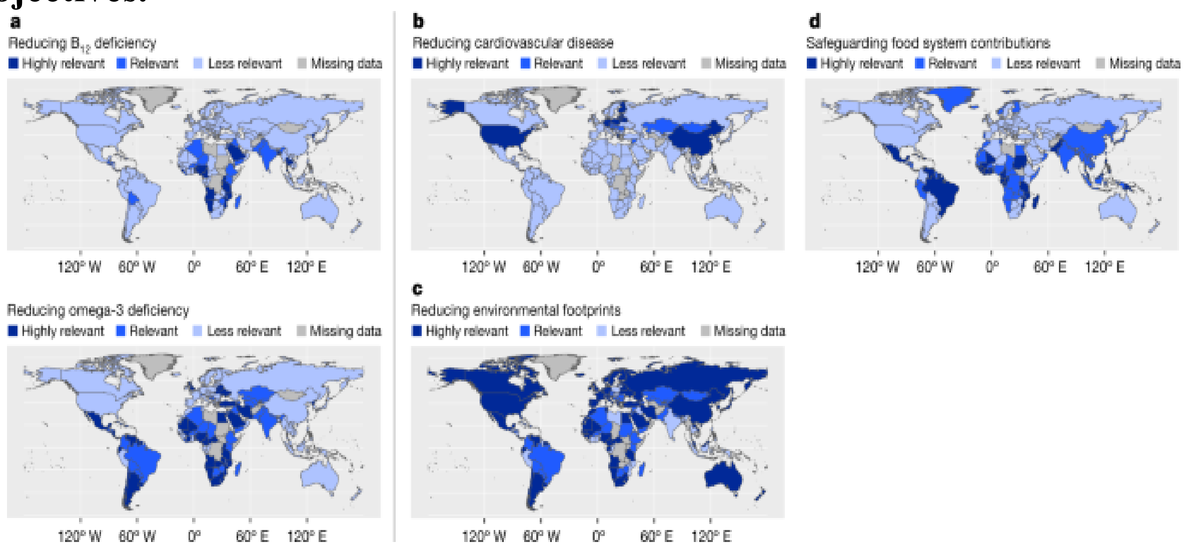
The potential contributions of blue foods to achieving food system ambitions depend on specific environmental, socio-economic and cultural contexts[60,61], which in turn are embedded in broader economic and political spheres[15]. We translate the blue food functions reviewed above into four policy objectives. These include leveraging consumption of blue food to: reduce vitamin B_{12} and omega-3 deficiencies; reduce non-communicable disease risks related to overconsumption of red meat, particularly cardiovascular disease; and reduce GHG consumption and production footprints. A fourth policy objective centres on safeguarding blue food contributions to nutrition, just economies, livelihoods and cultures under climate change. Each objective is mapped to individual country contexts on the basis of publicly available data (Table 1), to assess the broad relevance of each policy across nations. For example, using proxy variables for insufficient nutrient intake across populations (summary exposure values of vitamin B_{12} and omega-3 fatty acids), alongside blue food availability (through trade or domestic production), we identify countries for which reducing vitamin B_{12} or omega-3 deficiencies among nutritionally vulnerable populations is particularly relevant (see Supplementary Table 1 for details on all variables, underlying assumptions and cutoff values). Conditions for relevance were informed by key literature and expert assessment by the interdisciplinary pool of authors. This mapping is a first step towards a more context-specific articulation of the multi-dimensional policy relevance of blue foods in food systems around the world, and could be enhanced as further data at subnational level, or for small-scale operations, become available.

Table 1 Four policy objectives delineating the role blue foods can play in addressing social, environmental and nutritional challenges of food systems in different contexts

National relevance of policy objectives

The relevance of each blue food policy objective is mapped across nations globally (Fig. 1). The degree of relevance of each policy objective is evaluated by a set of rules and cutoff points (detailed in Supplementary Table 1), including a sensitivity analysis of cutoffs (Extended Data Figs. 2–6). An interactive website (https://gedb.shinyapps.io/BFA_synthesis/) presents all data and allows users to adjust cutoff points to explore the impacts of this on the relevance of each policy objective.

Fig. 1: National relevance of blue food in supporting four policy objectives.



Policy objective relevance is based on how well each nation matched the conditions for when blue foods could be expected to contribute to achieving food system ambitions (see Supplementary Table 2 for formalized inclusion criteria). **a**–**d**, The national relevance of the policies relating to reducing blue-food-sensitive deficiencies (vitamin B_{12} (top) and omega-3 (bottom); **a**), reducing the burden of cardiovascular disease (**b**), reducing environmental footprints of food consumption and production (**c**) and safeguarding blue food contributions under climate change (**d**). Readers can examine the detailed objectives matching individual countries, and explore effects of different cutoffs at https://gedb.shinyapps.io/BFA_synthesis/.

Fewer countries (43) were estimated to have >10% of their population at risk from inadequate intake of vitamin B_{12}, compared to omega-3 deficiency (89 countries), which predominantly affects African and South American nations. Many of these nations also have a high availability of blue foods, making them well positioned to address deficiencies by promoting access and facilitating consumption of culturally relevant blue food, especially among nutritionally vulnerable population segments (Fig. 1a).

Countries with red meat intake above the threshold recommended as environmentally sustainable and healthy[31], who also have a high incidence of cardiovascular disease, are primarily located in the global North, with the exception of several small-island states. In many of these countries, blue food is currently available (Fig. 1b and Supplementary Table 2). In such settings, moderate consumption of seafood with low environmental impact could be encouraged as a stepping stone away from high intake of red meat.

A substantial number of countries (124) also have a high intake of ruminant meat, contributing to high dietary GHG footprints (Fig. 1c). Many of these countries have blue food available because they are big importers (for example, Belgium) or big producers (for example, Chile and Norway), or they both produce and import (for example, France and Denmark). Although they may export some of their blue food at present, our mapping identifies countries that, with a shift in policy or prioritization, could retain some of their domestic production for domestic consumption. This would trade off export revenue, and highlights the need to balance several policy goals, discussed below.

At present, blue foods play an important role for nutrition, livelihoods or national revenue, in a substantial number of countries (103), particularly in the global South and among Indigenous communities across the global North[4]. Combining such findings with analysis of climate hazards identifies countries with high future risk, for whom climate adaptation of blue food systems will be particularly important (Fig. 1d). We illustrate some adaptation options below.

It is important to note that for a sizeable portion of countries, certain blue food policy objectives are less relevant, according to our analysis. This does not mean that food systems in these countries are devoid of challenges, but

blue foods are not a panacea and do not offer suitable means to improve food systems in all geographies at present.

Overlapping policy relevance

For some countries, several policy objectives are relevant. Figure 2 shows the degree of overlap between policy relevance, in terms of the number of countries for which two objectives are both relevant. Some policies show a high degree of overlap, indicating possible win-wins. For example, in most (75%) of the 89 countries for which omega-3-enhancing policies are relevant, reducing environmental footprints is also a relevant objective. Similarly, for most (82%) of the 22 countries dealing with high cardiovascular disease risk, promoting blue foods over red (particularly ruminant) meat overconsumption as part of a whole-diet approach would simultaneously address health and environmental concerns.

Fig. 2: Overlap in relevance between different policy objectives.

	Reducing blue-food-sensitive deficiencies		Reducing cardiovascular disease (n = 22)	Reducing environmental footprints (n = 124)	Safeguarding food system contributions (n = 103)
	Vitamin B_{12} (n = 43)	Omega-3 (n = 89)			
Vitamin B_{12}		44% (n = 39)	0% (n = 0)	19% (n = 23)	26% (n = 27)
Omega-3	91% (n = 39)		18% (n = 4)	54% (n = 67)	46% (n = 47)
Reducing cardiovascular disease	0% (n = 0)	4% (n = 4)		15% (n = 18)	3% (n = 3)
Reducing environmental footprints	53% (n = 23)	75% (n = 67)	82% (n = 18)		50% (n = 52)
Safeguarding food system contributions	63% (n = 27)	53% (n = 47)	14% (n = 3)	42% (n = 52)	

Overlap (%): 100, 75, 50, 25, 0

The numbers in parentheses in the top row represent the total number of countries for which each policy is relevant. Each cell shows the number of countries (in parentheses) for which both column- and row-heading policies are relevant, as a proportion of countries for which the column-heading

policy is relevant. Relevance in this figure indicates countries categorized as 'highly relevant' or 'relevant' for a given policy.

It is noteworthy that 91% of countries with vitamin B_{12} deficiencies also show high levels of omega-3 deficiency. Vitamin B_{12} deficiency seems to reflect more general undernutrition of the population, whereas omega-3 deficiency (specifically DHA and EPA) is caused by low intake of blue foods. The large overlap is thus explained by most of the 42 countries whose populations are at risk of malnourishment also lacking in consumption of blue food.

In 50% of the 103 countries for which blue foods play an important role for nutrition, livelihoods or revenue, blue foods could also represent an avenue to reduce environmental footprints of ruminant meat consumption. Furthermore, in 46% of these 103 countries reducing omega-3 deficiencies is also relevant, and is similarly reflected in the 53% overlap in relevance between countries with high omega-3 deficiency and safeguarding food system contributions. In these settings, policies that can reduce certain types of malnutrition by implementing climate adaptations that ensure access to low-environmental-impact blue foods, while also securing quality jobs (that is, welfare benefits) and removing barriers to wealth-generating benefits, could therefore have the potential to generate substantial co-benefits[17,62]. Below we explore how potential co-benefits flagged by Fig. 2 can be realized.

Harnessing diversity for co-benefits

Achieving globally agreed targets, such as zero hunger, good health, healthy aquatic and terrestrial environments, and a stable climate requires systems thinking[11,63,64]. Optimizing one policy domain often leads to both positive and negative spillover effects in several other sectors[65,66]. Addressing food system complexities that span land and sea, and/or encompass production, processing, trade and consumption, will be feasible only through systemic food policy[64,66] that identifies co-benefits between policy objectives and actions to achieve them[67]. A systemic food policy agenda can provide a clearer understanding of the potential of blue food diversity for navigating

trade-offs and realizing synergies between blue food policy objectives. Below we explore opportunities for co-benefits across the policy objectives proposed above. For each, we discuss how ensuring diversity in blue food actors and blue food performance across the domains of health, nutrition, environmental impact and climate risk can help realize synergistic system-level outcomes. Overall, we argue that in any setting, shaping or maintaining food environments that make blue foods an attractive food choice is a prerequisite for achieving the multiple food system goals highlighted by this paper.

Human health and environmental sustainability

Enabling this synergy will depend on the ability of sustainably sourced blue food to displace currently consumed foods with high environmental impact. Some aquatic foods, such as bivalves and small fish, are nutrient dense and have low environmental footprints[1,2] offering an environmentally sustainable way to address both vitamin B_{12} and omega-3 deficiencies and cardiovascular disease risk. Along with cultural preferences, smell and taste, safety concerns and eating habits[68,69], price is key for determining household consumption[42]. Access and affordability are therefore prerequisites for blue foods to reduce nutrient deficiencies, cardiovascular disease risk and dietary environmental footprints[70]. However, blue food diversity means that income is a poor predictor of consumption when relying on aggregate data categories such as ‘fish’[10,71]. At present, some blue foods are more expensive than other animal protein, particularly in developing contexts[63,72], but in many settings they represent affordable sources of key nutrients[10,63,73,74]. Increasing or protecting affordability will require the commoditization of low-environmental-impact blue foods through policy and regulation that promote sustainable intensification and supply chain transformation[71,75]. This can include public incentives for directing research and development investment towards specific species and production systems, and/or market incentives for value chain actors to reorient trade to low-income and nutritionally vulnerable consumers, and prioritizing increased nutrition over growth in production volumes and monetary value.

Although consumption of blue food is projected to increase (about 80% in edible weight by 2050 assuming constant prices and balance between supply

and demand[10]), the resulting nutrition and environmental impacts will depend on the substitutability among blue foods and other animal-source foods in national diets. Substituting all red meats for blue foods is neither feasible nor desirable, and adding or increasing animal-source blue foods to diets of wealthy consumers, already rich in animal-source foods, would fundamentally undermine the role of blue foods in delivering healthier and less environmentally harmful dietary outcomes[2]. Fish–meat substitutability has not been widely studied, but the possibility of replacing meats with blue foods or plant-based alternatives seems to be an attractive policy option[35,72]. Strategies to achieve these goals could include combining soft policy tools such as dietary guidelines or behavioural nudging to mainstream eating and cooking blue foods[76,77], with harder regulatory interventions and economic disincentives for high-carbon-emissions food[78,79,80,81].

Livelihoods, economies, health and environmental sustainability

Investments in blue food innovations have the potential to yield inclusive livelihoods and systems that produce nutritious, affordable and environmentally sustainable blue foods[14,75]. Such synergies again depend on which species and production modes are pursued, their variable environmental performance and nutrient density, and what barriers to access exist[1,2,82]. Production modes also vary greatly[14], from un-mechanized small-scale fisheries and farming to industrial-scale, highly specialized operations. These different production modes, and the power dynamics of supply chains developed for distribution, generally affect their contribution to equitable wealth and welfare distribution[14,17]. Policy levers are therefore needed that can improve equity by removing barriers to wealth-generating benefits. This can entail inclusive financing, infrastructure and governance that lends voice and rights to all actors and avoids displacement by competitive sectors, but also maintaining traditional access rights to nutritious blue foods; all as part of efforts to implement the human right to food[14,17,62,83]. Improving equity can also yield further benefits. For instance, increasing gender equity has been found to also improve nutritional outcomes for families[84].

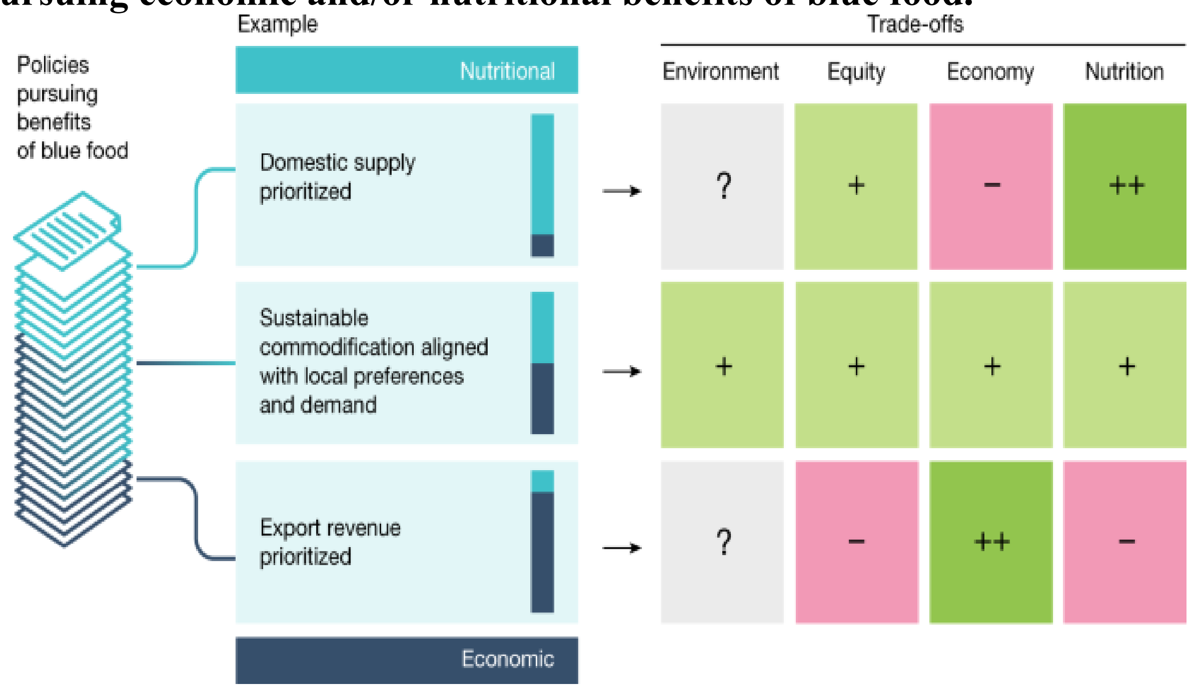
Climate resilience and blue food production, employment or revenue

Climate change will affect all aspects of aquatic food systems, from production to consumption, and threatens to undermine their contribution to the health, economies, culture and livelihoods of billions of people[20]. The substantial contribution that blue foods already make, particularly for livelihoods and diets, in many nations[85,86] (Fig. 1d) underscores the importance of strengthening blue food system resilience as no- or low-regret adaptation options[87]. Climate-smart production, supported through finance and adaptive governance, can reduce future disruption by promoting a multitude of different blue foods, and thus also take advantage of new opportunities that come with changing species and conditions. Examples include farming several thermally tolerant species, or introducing more flexible catch guidelines to cater for geographically extended species ranges and migration patterns of fish and fishers[88]. Addressing the current unsustainability of many fisheries by regulating harvestable quantities would simultaneously enhance stock resilience through maintenance of higher genetic diversity and thus adaptive capacity. Larger stocks are also less likely to crash when exposed to periodic shocks, such as El Niño and marine heatwaves[89]. For aquaculture, relying on a diversity of species could provide response diversity across a number of critical dimensions such as temperature, salinity or oxygen. While increasing the diversity of species to reduce climate sensitivity and increase adaptive capacity, aquaculture could also reduce the focus on fed species and promote the development of non-fed production systems[90]. Additionally, by valuing the diversity of skills and knowledge encompassed by small-scale actors and enabling their capacities to innovate and adapt to changing environmental and economic conditions, nations could further invest in the resilience of their aquatic food system (S.R.B., manuscript in preparation)[14]. Enhanced capabilities of the small-scale sector would also increase their ability to establish rights over resources, promoting more equitable forms of production and employment. Finally, disincentivizing high concentration of economic power in supply chains, characterized by the singular pursuit of efficiency gains, and mechanization at the expense of jobs[14,54] will be important to ensure that potential synergies between SSFA diversity, climate resilience and equity materialize.

Navigating trade-offs

The complex nature of food systems, including aquatic ones, means that any action to improve performance along some dimensions will trade off performance on one or several others[65]. We identify and elaborate on three bundles of substantial trade-offs that need to be considered, but which can be navigated and minimized by making strategic use of the diversity of blue food species and production systems. We visualize an example of such trade-offs using the pursuit of either economic or nutritional blue food benefits through domestic consumption or export (Fig. 3).

Fig. 3: Example of hypothetical trade-offs associated with policies pursuing economic and/or nutritional benefits of blue food.



The figure illustrates one set of trade-offs in policy outcomes that may result across the dimensions of environment, equity, economy and nutrition, depending on the degree of prioritization of either increasing domestic blue food supplies for nutritional outcome, or maximizing monetary value through exports of blue foods. The degree of emphasis placed on either policy goal is represented by the blue bars. Likely outcomes for each dimension are represented by coloured boxes and the strength of outcome is represented by plus and minus symbols; with positive outcomes depicted in green, and negative in pink. Sustainable commodification aligned with local preferences and demand represents an example of how a balance could be struck to optimize positive environmental, inclusive, economic and

nutritional outcomes. Unknown impacts, or where policy objectives are judged to not have a strong impact, are depicted in grey. E. Wikander/Azote.

Environmental sustainability versus nutritional content of aquaculture products

In aquaculture, a pressing challenge has been to reduce reliance on wild fish for feed[50,91,92] by incorporating plant-based ingredients and recycled animal processing wastes in feeds[18,93]. However, such feeds may compromise the nutritional value of the fish produced[94] and divert produce that could be used for direct human consumption. Continued innovation to develop alternative feeds that combine lower environmental footprint with high nutritional quality will therefore be important, along with lowering feed conversion ratios[2,92], but the latter will pit improved local environmental performance against higher-quality resource requirements with consequences for sustainable and ethical resource management. Policies that promote supportive structures of governance, infrastructure and financial access to new technologies and high-quality feeds for small-scale producers will contribute to mainstreaming a move away from wild fish feeds. A regulatory environment tailored to the specific needs of blue foods, as opposed to outdated and agriculturally focused rules such as bans on use of non-ruminant processed animal proteins and genetically modified organisms, can enable rather than hinder inclusion of new protein sources. It could also avoid perverse impacts and enhance regulatory coherence for improved innovation, market access and sustainability[51].

Domestic consumption versus export revenues

Production of blue foods for export, and allocation of fishing rights to foreign fleets, offer economic opportunities for governments, individual businesses and fishers. However, these actions can undermine domestic consumption and local livelihoods[95,96] if fishers and farmers are displaced from productive fishing or aquaculture areas[97], or suffer knock-on impacts of damaging industrial practices and overexploitation. Small-scale producers often face a tension between local needs[98] and connecting to export markets with higher profits that leave them vulnerable to global power dynamics,

price fluctuations and supply chain disruptions[99]. In some cases, such as Chile, rising blue food export is associated with declining national consumption in favour of terrestrial meat[10]. As demand for blue foods rises among affluent population groups because of their contributions to health and reduced environmental footprint, prices will probably increase, exacerbating this tension. Implementing environmentally sustainable commodification will therefore be important, but must ensure that small-scale actors are not marginalized in the process (for example, ref. [100]), This requires policies that encourage collaborative practices across production scales. Cooperatives and coalitions can support complementary and synergistic production and resource access across producers[101], and inclusive jurisdictional and landscape approaches offer means to reconcile the diverse incentives and capabilities of actors in blue food production systems, while addressing ecological and geographical mismatches of current ratings and certification systems[102]. Exempting some domestic production from export is another way to secure food access but must align with local preferences, ensuring demand exists and producers are not disenfranchised[103].

Efficiency, affordability and availability versus diversification and resilience

The global capacity to produce increased quantities of nutrient-dense, yet low-impact aquatic foods will influence the severity of trade-offs that emerge between blue food policy objectives. Commodification often offers efficiency and economies of scale, making blue foods more affordable and accessible[75], but may compromise nutrition, squeeze out small producers and processors or outcompete them in markets if measures are not in place to safeguard their livelihoods[104]. For example, large-scale production of tilapia and pangasius offers inexpensive sources of aquatic foods, but in some markets they have replaced more nutritious indigenous fish[82]. Ultimately, efficiencies must be balanced against food sovereignty and the many contributions of blue foods, distinct from their monetary value[17]. Policies to retain or enhance the diversity of blue food production modes, actors and species are essential for the capacity of nations and regions to build resilience against shocks associated with, for example, climate

change[20,105], trade[19,57] or new diseases[106]. Examples include government support funds to provide financial relief for small businesses highly vulnerable to environmental and trade fluctuations[106], and improved accessibility to production-related insurance[107]. The combination of species and productions systems that provide most resilience to a changing climate will be highly context specific, yet the species that offer opportunities for efficiencies and bulk production under a changing climate (such as tilapia with a high temperature tolerance range) may not be the most culturally appropriate or nutrient-dense aquatic foods[1]. Navigating this apparent trade-off could involve complementing bulk production of fewer species with environmentally sustainable cultivation and capture of a diversity of species that provide additional nutrition (for example, dried fish powder) and more inclusive supply chains[77].

Policy ambitions need bold visions

The coronavirus disease 2019 pandemic has changed many aspects of our lives: how we work, travel and eat. For better or worse, it has shown that radical change is feasible in a short amount of time. For example, the pandemic highlighted how the small-scale blue foods sector was able to convey resilience and fill nutritional gaps left by interrupted global markets in some contexts, whereas in others it was left highly vulnerable[14,106]. Such shocks illustrate that backcasting is not the only, or the best, way to understand the future. Envisaging alternative futures (for example, through scenarios) may be instrumental for altering entrenched, unhealthy and unsustainable ways of producing and consuming food[80,91].

We have outlined four roles that blue foods can play now and in the future and have translated these into broad policy objectives that—if actions are developed to achieve them—could contribute to achieving articulated food system ambitions (for example, United Nations Food Systems Summit 2021). Our analysis shows that the health, environmental, economic and welfare benefits that nations derive from blue foods are diverse[60,61]. We therefore provide an analytical framework and an interactive tool (https://gedb.shinyapps.io/BFA_synthesis/) for decision makers to explore

how this diversity affects the relevance of specific blue food objectives in specific contexts.

However, regardless of how environmentally sustainably produced blue foods are, the global demand for blue foods to address disease and environmental impact in one set of countries (Fig. 1b,c) may reduce the availability and affordability of blue foods for achieving improved nutritional status for vulnerable populations in another (Fig. 1a). Governments can address these tensions by regulating trade and by ensuring that diets incorporating blue foods are considered alongside other means of achieving environmentally sustainable and healthy food system outcomes, such as various forms of more diverse and plant-rich diets[31,43,72,81]. Decisions regarding the role that blue foods can and should play for any nation's journey towards a more nutritious, equitable and less environmentally harmful food system therefore need to be grounded in local context and availability of aquatic foods, but also availability and affordability of a diversity of alternatives that are equally healthy and sustainable. Furthermore, the dietary shifts associated with the nutrition transition[34] are neither globally universal, nor inevitable. Despite their growing incomes, India and most countries in Asia–Pacific, much of the Middle East and some Latin American nations show low terrestrial meat preferences, with a higher share of protein coming from other sources, such as legumes and seafood[10]. Nations whose diets were previously constrained by low income are therefore now well placed to lead the way to sustainable and healthy eating.

Methods

Assessing degree of policy relevance

To assess the degree of relevance of each policy for each country, we rely on theory and expert-guided typology building. Such an approach centres on classifying countries on the basis of a set of a priori assumptions about the conditions when blue food policies are relevant. The analysis has three steps.

Step one uses theory and expert assessments to build a data table of conditions that logically explain the relevance or non-relevance of each of

the four policies. Conditions are variables that explain an outcome. In our analysis, these variables represent proxy variables that logically explain the relevance or non-relevance of the four policies in focus (Supplementary Table 1). The proxy variables correspond to national averages of publicly available (or published) datasets. Following best practice for related methods, such as qualitative comparative analysis[109], thresholds for inclusion (that is, cutoffs for when a policy objective is considered relevant on the basis of a country's statistic) were set on the basis of theoretical knowledge where available (Supplementary Table 1 and Extended Data Fig. 1). For many variables, however, no theoretically established cutoff existed. Thresholds were then set on the basis of natural breaks in the data after deliberation with authors of relevant expertise (see Extended Data Figs. 2–6). For all cutoff values, we provide a transparent justification for the selection and specify the type of disciplinary expertise leveraged to assess cutoffs for each variable (Supplementary Table 1).

A second step involves developing Boolean logic solution formulae that allow us to classify countries in relation to the outcome variable 'degree of policy relevance' (highly relevant, relevant, less relevant and missing data). We use a crisp set methodology to define which countries are relevant for a particular policy (see Supplementary Information for elaborated justification). Crisp sets assign cases (countries) a binary value for each variable. The binary value is based on whether the data for the country fall above or below the pre-determined cutoff (Supplementary Table 1). Logic solution formulae specify the combination of binary conditions that results in a given level of policy relevance (Supplementary Table 2), and these solution formulae are based on expert judgement and logic (Supplementary Table 1). The combination of AND and OR statements in the solution formulae highlight the distinction between necessary and sufficient conditions, akin to how these are conceptualized in qualitative comparative analysis[110].

All datasets used as input into the Boolean analysis (referenced in Supplementary Table 1) are freely available through peer-reviewed publications or publicly available databases. These include: ref. [1]; Food and Agriculture Organization of the United Nations (FAO) Fishery and Aquaculture Statistics, Global capture production 1950–2019 (FishstatJ),

available at https://www.fao.org/fishery/en/statistics (ref. [13]); global expanded nutrient supply model, available at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/5LC3SI ; World Health Organization, Global health estimates: leading causes of DALYs, available at https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys; (ref. [15]); FAO Yearbook, Fishery and Aquaculture Statistics 2018, available at https://www.fao.org/fishery/en/publications/269665; ILOSTAT labour statistics (2020), available at https://ilostat.ilo.org/; World Development Indicators (World Bank) DataBank (2012), available at https://databank.worldbank.org/reports.aspx?source=world-development-indicators (ref. [37]); FAOSTAT Food Balances, available at http://www.fao.org/faostat/en/#data/FBS; the variable 'hazard by system' in the web application uses data presented in the extended data for ref. [20], available at https://doi.org/10.1038/s43016-021-00368-9.

One key reason for choosing crisp set methodology was that we wanted to maximize the ease of interpretation and potential use. Crisp sets arguably retain less information richness than fuzzy sets (for which membership of cases is not binary but assigned as degrees of membership to different categories). However, although partial set membership allows for more information from the underlying data to be maintained, it is also likely to result in situations of partial relevance in the outcome variable (degree of policy relevance). In other words, one could easily end up in a situation in which a country is classified as 33% relevant. This would be exceedingly hard for readers to interpret and act on. In other words, crisp sets were chosen in order for countries to receive a clear classification of relevance (highly relevant, relevant, less relevant and missing data) in our analysis. Another reason for opting for crisp sets is the above noted lack of scientific consensus to guide the exact cutoffs for all variables assessing the conditions. Fuzzy set analysis requires several such decisions to be made as each variable is divided into a minimum of three sets (as opposed to a case simply being in the set = 1, or out = 0), and would have thus increased the uncertainty of the analysis. We recognize that even with our analysis some cutoffs could be up for discussion. We therefore invite readers to explore different threshold values and see the change in outcome in the web-based

tool available at https://gedb.shinyapps.io/BFA_synthesis/. This tool also means that, as scientific evidence for a particular cutoff value becomes available or updated, this information can easily be applied to revise the classification of nations.

The third step of our analytical approach involves matching the set configurations in the data table (step 1) to the Boolean logic solution formulae designed in step 2, to assign each case (country) to the outcome variable ‘degree of policy relevance’ (Supplementary Table 2). This outcome variable (for each policy objective) forms the results presented in Fig. 1, and forms the basis of the overlap analysis presented in Fig. 2.

Sensitivity analysis

To assess the sensitivity of our results to variations in thresholds, we opted for a one-at-a-time sensitivity approach[111] (Extended Data Figs. 2–6) owing to the simplicity of Boolean rules used in this analysis. This was combined with a visual examination of the underlying distribution of each variable used in the analysis, in relation to our set threshold (Extended Data Fig. 1). The typology classification model was re-run changing one variable threshold at a time, leaving all else constant, and assessing the change in the number of countries in each outcome category (‘highly relevant’, ‘relevant’ and ‘less relevant’). The threshold was varied across the full range of the variable. The sensitivity analysis highlights to what degree the underlying distribution of the data (Extended Data Fig. 1), as well as the Boolean logic on which the classification model is based (Supplementary Table 2), influences the sensitivity of the threshold.

Assessing overlap in policy relevance

To assess overlap in policy objective relevance among nations, we conducted pairwise comparison of policy objectives. We calculated the percentage of the set of countries to whom a specific policy was deemed relevant, and that was also deemed relevant for a second policy objective. Nations classified as ‘highly relevant’ or ‘relevant’ were combined for the purpose of this analysis, and countries with missing data for either policy were not considered.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated and analysed during the study are available in the Stockholm University Library Dataverse (https://doi.org/10.7910/DVN/ILA0XI).

Code availability

R code used for the Boolean analysis and sensitivity analysis, as well as for producing the figures and web application, is available at https://github.com/emmywas/BFA_Policy_analysis. All coding was carried out in R (version 4.2.0).

References

1. Golden, C. D. et al. Aquatic foods for nourishing nations. *Nature* **598**, 315–320 (2021).

2. Gephart, J. A. et al. Environmental performance of blue foods. *Nature* **597**, 360–365 (2021).

3. *Continuous Update Project Expert Report 2018. Meat, Fish and Dairy Products and the Risk of Cancer* (World Cancer Research Fund/American Institute for Cancer Research, 2018).

4. Cisneros-Montemayor, A. M., Pauly, D., Weatherdon, L. V. & Ota, Y. A global estimate of seafood consumption by coastal Indigenous peoples. *PLoS ONE* **11**, e0166681 (2016).

5. Halpern, B. S. et al. Opinion: Putting all foods on the same table: achieving sustainable food systems requires full accounting. *Proc. Natl*

Acad. Sci. USA **116**, 18152–18156 (2019).

6. Bennett, A. et al. Recognize fish as food in policy discourse and development funding. *Ambio* **50**, 981–989 (2021).

7. Béné, C. et al. Feeding 9 billion by 2050 – putting fish back on the menu. *Food Secur.* **7**, 261–274 (2015).

8. Koehn, J. Z., Allison, E. H., Golden, C. D. & Hilborn, R. The role of seafood in sustainable diets. *Environ. Res. Lett.* **17**, 035003 (2022).

9. Tlusty, M. F. et al. Reframing the sustainable seafood narrative. *Glob. Environ. Change* **59**, 101991 (2019).

10. Naylor, R. L. et al. Blue food demand across geographic and temporal scales. *Nat. Commun.* **12**, 5413 (2021).

11. Farmery, A. K. et al. Blind spots in visions of a 'blue economy' could undermine the ocean's contribution to eliminating hunger and malnutrition. *One Earth* **4**, 28–38 (2021).

12. Koehn, J. Z. et al. Fishing for health: do the world's national policies for fisheries and aquaculture align with those for nutrition? *Fish Fish.* https://doi.org/10.1111/FAF.12603 (2021).

13. *Fishery and Aquaculture Statistics. Global Capture Production 1950–2019 (FishstatJ)* https://www.fao.org/fishery/en/statistics (FAO, 2021).

14. Short, R. E. et al. Harnessing the diversity of small-scale actors is key to the future of aquatic food systems. *Nat. Food* **2**, 733–741 (2021).

15. Teh, L. C. L. & Sumaila, U. R. Contribution of marine fisheries to worldwide employment. *Fish Fish.* **14**, 77–88 (2013).

16. Loring, P. A. et al. in *Transdisciplinarity for Small-Scale Fisheries Governance: Analysis and Practice* (eds Chuenpagdee, R. & Jentoft, S.) 55–73 (Springer, 2019).

17. Hicks, C. C. et al. Rights and representation support justice across aquatic food systems. *Nat. Food* **3**, 851–861 (2022).

18. Troell, M. et al. Does aquaculture add resilience to the global food system? *Proc. Natl Acad. Sci. USA* **111**, 13257–13263 (2014).

19. Ferguson, C. E. et al. Local practices and production confer resilience to rural Pacific food systems during the COVID-19 pandemic. *Mar. Policy* **137**, 104954 (2022).

20. Tigchelaar, M. et al. Compound climate risks threaten aquatic food system benefits. *Nat. Food* **2**, 673–682 (2021).

21. *Nutrition and Food Systems. A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security* (Food and Agricultural Organization, 2017).

22. von Braun, J., Afsana, K., Fresco, L. O. & Hassan, M. Food systems: seven priorities to end hunger and protect the planet. *Nature* **597**, 28–30 (2021).

23. *Transforming Our World: the 2030 Agenda for Sustainable Development* (United Nations, 2015).

24. Sumaila, U. R. & Tai, T. C. End overfishing and increase the resilience of the ocean to climate change. *Front. Mar. Sci.* **7**, 523 (2020).

25. *The State of World Fisheries and Aquaculture 2020. Sustainability in Action* (FAO, 2020).

26. Thilsted, S. H. et al. Sustaining healthy diets: the role of capture fisheries and aquaculture for improving nutrition in the post-2015 era. *Food Policy* **61**, 126–131 (2016).

27. Roos, N., Wahab, M. A., Chamnan, C. & Thilsted, S. H. The role of fish in food-based strategies to combat vitamin A and mineral deficiencies in developing countries. *J. Nutr.* **137**, 1106–1109 (2007).

28. Starling, P., Charlton, K., McMahon, A. T. & Lucas, C. Fish intake during pregnancy and foetal neurodevelopment—a systematic review of the evidence. *Nutrients* **7**, 2001–2014 (2015).

29. Byrd, K. A., Pincus, L., Pasqualino, M. M., Muzofa, F. & Cole, S. M. Dried small fish provide nutrient densities important for the first 1000 days. *Matern. Child Nutr.* **17**, e13192 (2021).

30. Starling, P., Charlton, K., McMahon, A. T. & Lucas, C. Fish intake during pregnancy and foetal neurodevelopment—a systematic review of the evidence. *Nutrients* **7**, 2001–2014 (2015).

31. Willett, W. et al. Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet* **393**, 447–492 (2019).

32. Wolk, A. Potential health hazards of eating red meat. *J. Intern. Med.* **281**, 106–122 (2017).

33. Richi, E. B. et al. Health risks associated with meat consumption: a review of epidemiological studies. *Int. J. Vitam. Nutr. Res.* **85**, 70–78 (2015).

34. Popkin, B. M. & Gordon-Larsen, P. The nutrition transition: worldwide obesity dynamics and their determinants. *Int. J. Obes.* **28**, S2–S9 (2004).

35. Zeng, L. et al. Trends in processed meat, unprocessed red meat, poultry, and fish consumption in the United States, 1999–2016. *J. Acad. Nutr. Diet.* **119**, 1085–1098 (2019).

36. *OECD-FAO Agricultural Outlook 2012-2021* (OECD, accessed 31 August 2021); https://stats.oecd.org/Index.aspx?DataSetCode=HIGH_AGLINK_2012.

37. *World Development Indicators* https://databank.worldbank.org/reports.aspx?source=world-development-indicators (The World Bank, 2012).

38. Manson, J. E. et al. Marine n−3 fatty acids and prevention of cardiovascular disease and cancer. *N. Engl. J. Med.* **380**, 23–32 (2019).

39. Miller, V. et al. Evaluation of the quality of evidence of the association of foods and nutrients with cardiovascular disease and diabetes: a systematic review. *JAMA Netw. Open* **5**, e2146705(2022).

40. Abdelhamid, A. S. et al. Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst. Rev.* **11**, CD003177 (2018).

41. Guasch-Ferré, M. et al. Meta-analysis of randomized controlled trials of red meat consumption in comparison with various comparison diets on cardiovascular risk factors. *Circulation* **139**, 1828–1845 (2019).

42. Moberg, E. et al. Combined innovations in public policy, the private sector and culture can drive sustainability transitions in food systems. *Nat. Food* **2**, 282–290 (2021).

43. Poore, J. & Nemecek, T. Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2018).

44. Parker, R. W. R. et al. Fuel use and greenhouse gas emissions of world fisheries. *Nat. Clim. Change* https://doi.org/10.1038/s41558-018-0117-x (2018).

45. Ainsworth, R., Cowx, I. G. & Funge-Smith, S. A review of major river basins and large lakes relevant to inland fisheries. *FAO Fish. Aquac. Circ.* https://doi.org/10.4060/CB2827EN (2021).

46. Hilborn, R., Banobi, J., Hall, S. J., Pucylowski, T. & Walsworth, T. E. The environmental cost of animal source foods. *Front. Ecol. Environ.* **16**, 329–335 (2018).

47. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).

48. Pérez Roda, M. A. et al. *A Third Assessment of Global Marine Fisheries Discards* (FAO, 2019).

49. Robinson, J. P. W. et al. Managing fisheries for maximum nutrient yield. *Fish Fish.* https://doi.org/10.1111/FAF.12649 (2022).

50. Naylor, R. L. et al. A 20-year retrospective review of global aquaculture. *Nature* **591**, 551–563 (2021).

51. Henriksson, P. G. et al. Interventions for improving the productivity and environmental performance of global aquaculture for future food security. *One Earth* **4**, 1220–1232 (2021).

52. Hamilton-Hart, N. & Stringer, C. Upgrading and exploitation in the fishing industry: contributions of value chain analysis. *Mar. Policy* **63**, 166–171 (2016).

53. Weeratunge, N. et al. Small-scale fisheries through the wellbeing lens. *Fish Fish.* **15**, 255–279 (2014).

54. Clapp, J. The problem with growing corporate concentration and power in the global food system. *Nat. Food* **2**, 404–408 (2021).

55. Hanich, Q. et al. Small-scale fisheries under climate change in the Pacific Islands region. *Mar. Policy* **88**, 279–284 (2018).

56. Selig, E. R. et al. Mapping global human dependence on marine ecosystems. *Conserv. Lett.* **12**, e12617 (2019).

57. Gephart, J. A., Deutsch, L., Pace, M. L., Troell, M. & Seekell, D. A. Shocks to fish production: identification, trends, and consequences. *Glob. Environ. Change* **42**, 24–32 (2017).

58. Allison, E. H., Béné, C. & Andrew, N. L. in *Small-Scale Fisheries Management: Frameworks and Approaches for the Developing World* (eds Pomeroy, R. & Andrew, N. L.) 206–238 (CABI, 2011).

59. Leslie, P. & McCabe, J. T. Response diversity and resilience in social-ecological systems. *Curr. Anthropol.* **54**, 114–143 (2013).

60. Ota, Y., Allison, E. H. & Fabinyi, M. Evolving the narrative for protecting a rapidly changing ocean, post-COVID-19. *Aquat. Conserv.*

Mar. Freshw. Ecosyst. **31**, 1925–1926 (2021).

61. Hulme, M. One Earth, many futures, no destination. *One Earth* **2**, 309–311 (2020).

62. Song, A. M. & Soliman, A. Situating human rights in the context of fishing rights – contributions and contradictions. *Mar. Policy* **103**, 19–26 (2019).

63. Herforth, A. et al. *Cost and Affordability of Healthy Diets across and within Countries: Background Paper for the State of Food Security and Nutrition in the World 2020* FAO Agricultural Development Economics Technical Study 309369 (FAO, 2020).

64. Imagine a world without hunger, then make it happen with systems thinking. *Nature* **577**, 293–294 (2020).

65. Herrero, M. et al. Articulating the effect of food systems innovation on the Sustainable Development Goals. *Lancet Planet. Heal.* **5**, e50–e62 (2021).

66. Tezzo, X., Bush, S. R., Oosterveer, P. & Belton, B. Food system perspective on fisheries and aquaculture development in Asia. *Agric. Human Values* **38**, 73–90 (2021).

67. Clark, M. A., Springmann, M., Hill, J. & Tilman, D. Multiple health and environmental impacts of foods. *Proc. Natl Acad. Sci. USA* **116**, 23357–23362 (2019).

68. Dubois, P., Griffith, R. & Nevo, A. Do prices and attributes explain international differences in food purchases. *Am. Econ. Rev.* **104**, 832–867 (2014).

69. Blake, C. E. et al. Elaborating the science of food choice for rapidly changing food systems in low-and middle-income countries. *Glob. Food Sec.* **28**, 100503 (2021).

70. Dey, M. M. et al. Demand for fish in Asia: a cross-country analysis. *Aust. J. Agric. Resour. Econ.* **52**, 321–338 (2008).

71. Gallet, C. The demand for fish: a meta-analysis of the own-price elasticity. *Aquac. Econ. Manag.* **13**, 235–245 (2009).

72. Springmann, M. et al. Health and nutritional aspects of sustainable diet strategies and their association with environmental impacts: a global modelling analysis with country-level detail. *Lancet Planet. Heal.* **2**, E451–E461 (2018).

73. Ryckman, T., Beal, T., Nordhagen, S., Chimanya, K. & Matji, J. Affordability of nutritious foods for complementary feeding in Eastern and Southern Africa. *Nutr. Rev.* **79**, 35–51 (2021).

74. Ryckman, T., Beal, T., Nordhagen, S., Murira, Z. & Torlesse, H. Affordability of nutritious foods for complementary feeding in South Asia. *Nutr. Rev.* **79**, 52–68 (2021).

75. Belton, B., Reardon, T. & Zilberman, D. Sustainable commoditization of seafood. *Nat. Sustain.* **3**, 677–684 (2020).

76. van Putten, I. et al. Fresh eyes on an old issue: demand-side barriers to a discard problem. *Fish. Res.* **209**, 14–23 (2019).

77. Koehn, J., Quinn, E. L., Otten, J. J., Allison, E. H. & Anderson, C. M. Making seafood accessible to low-income and nutritionally vulnerable populations on the U.S. West Coast. *J. Agric. Food Syst. Community Dev.* **10**, 171–189 (2020).

78. Röös, E. et al. *Policy Options for Sustainable Food Consumption: Review and Recommendations for Sweden* (Mistra Sustainable Consumption Project, 2021).

79. Fesenfeld, L. P., Wicki, M., Sun, Y. & Bernauer, T. Policy packaging can make food system transformation feasible. *Nat. Food* **1**, 173–182 (2020).

80. Farmery, A. K. et al. Food for all: designing sustainable and secure future seafood systems. *Rev. Fish Biol. Fish.* **32**, 101–121 (2021).

81. Springmann, M. et al. The healthiness and sustainability of national and global food based dietary guidelines: modelling study. *Br. Med. J.* **370**, m2322 (2020).

82. Bogard, J. R. et al. Higher fish but lower micronutrient intakes: temporal changes in fish consumption from capture fisheries and aquaculture in Bangladesh. *PLoS ONE* **12**, e0175098 (2017).

83. Allison, E. H. et al. Rights-based fisheries governance: from fishing rights to human rights. *Fish Fish.* **13**, 14–29 (2012).

84. Cole, S. M. et al. Gender accommodative versus transformative approaches: a comparative assessment within a post-harvest fish loss reduction intervention. *Gend. Technol. Dev.* **24**, 48–65 (2020).

85. Golden, C. D. et al. Nutrition: fall in fish catch threatens human health. *Nature* **534**, 317–320 (2016).

86. Lam, V. W. Y. et al. Climate change, tropical fisheries and prospects for sustainable development. *Nat. Rev. Earth Environ.* **1**, 440–454 (2020).

87. Heltberg, R., Siegel, P. B. & Jorgensen, S. L. Addressing human vulnerability to climate change: toward a ‘no-regrets’ approach. *Glob. Environ. Change* **19**, 89–99 (2009).

88. Clarke, T. M. et al. Climate change impacts on living marine resources in the Eastern Tropical Pacific. *Divers. Distrib.* **27**, 65–81 (2021).

89. Cheung, W. W. L., Jones, M. C., Reygondeau, G. & Frölicher, T. L. Opportunities for climate-risk reduction through effective fisheries management. *Glob. Change Biol.* **24**, 5149–5163 (2018).

90. Oyinlola, M. A. et al. Projecting global mariculture production and adaptation pathways under climate change. *Glob. Change Biol.* **28**, 1315–1331 (2022).

91. Gephart, J. A. et al. Scenarios for global aquaculture and its role in human nutrition. *Rev. Fish. Sci. Aquac.* **29**, 122–138 (2020).

92. Hua, K. et al. The future of aquatic protein: implications for protein sources in aquaculture diets. *One Earth* **1**, 316–329 (2019).

93. Shepon, A. et al. Sustainable optimization of global aquatic omega-3 supply chain could substantially narrow the nutrient gap. *Resour. Conserv. Recycl.* **181**, 106260 (2022).

94. Sprague, M., Dick, J. R. & Tocher, D. R. Impact of sustainable feeds on omega-3 long-chain fatty acid levels in farmed Atlantic salmon, 2006–2015. *Sci. Rep.* **6**, 21892 (2016).

95. Béné, C., Hersoug, B. & Allison, E. H. Not by rent alone: analysing the pro-poor functions of small-scale fisheries in developing countries. *Dev. Policy Rev.* **28**, 325–358 (2010).

96. Hicks, C. C. et al. Harnessing global fisheries to tackle micronutrient deficiencies. *Nature* **574**, 95–98 (2019).

97. Cohen, P. J. et al. Securing a just space for small-scale fisheries in the blue economy. *Front. Mar. Sci.* **6**, 171 (2019).

98. Kent, G. *Fish Trade, Food Security and the Human Right to Adequate Food* (FAO, 2003).

99. Stoll, J. S., Crona, B. I., Fabinyi, M. & Farr, E. R. Seafood trade routes for lobster obscure teleconnected vulnerabilities. *Front. Mar. Sci.* **5**, 239 (2018).

100. Mamun, A. A. et al. Export-driven, extensive coastal aquaculture can benefit nutritionally vulnerable people. *Front. Sustain. Food Syst.* **5**, 713140 (2021).

101. Mccay, B. J. et al. Cooperatives, concessions, and co-management on the Pacific coast of Mexico. *Mar. Policy* **44**, 49–59 (2014).

102. Kittinger, J. N. et al. Applying a jurisdictional approach to support sustainable seafood. *Conserv. Sci. Pract.* **3**, e386 (2021).

103. Fakhri, M. *The Right to Food in the Context of International Trade Law and Policy.* Note by the Secretary, United Nations General Assembly, distributed on 22 July 2020 (UN, 2020).

104. Arthur, R. I. et al. Small-scale fisheries and local food systems: transformations, threats and opportunities. *Fish Fish.* https://doi.org/10.1111/FAF.12602 (2021).

105. Barange, M. et al. Impacts of climate change on marine ecosystem production in societies dependent on fisheries. *Nat. Clim. Change* **4**, 211–216 (2014).

106. Belton, B. et al. COVID-19 impacts and adaptations in Asia and Africa's aquatic food value chains. *Mar. Policy* **129**, 104523 (2021).

107. van Anrooy, R. *Review of the Current State of World Aquaculture Insurance* (Food and Agriculture Organization of the United Nations, 2006).

108. Tlusty, M. F., Hardy, R. & Cross, S. F. Limiting size of fish fillets at the center of the plate improves the sustainability of aquaculture production. *Sustainability* **3**, 957–964 (2011).

109. Rihoux, B. & De Meur, G. in *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (eds Rihoux, B. & Ragin, C. C.) 33–68 (Sage, 2009).

110. Rihoux, B. & Ragin, C. (eds.) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques* (Sage, 2009).

111. Hamby, D. M. A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Monit. Assess.* **32**, 135–154 (1994).

Acknowledgements

This paper is part of the Blue Food Assessment (https://www.bluefood.earth), a comprehensive examination of the role of

aquatic foods in building healthy, sustainable and equitable food systems. The assessment was supported by the Builders Initiative, the MAVA Foundation, the Oak Foundation and the Walton Family Foundation. We thank all scientific leaders of the Blue Food Assessment for their intellectual input on this paper, as well as S. Maniatakou and G. Parlato for research assistance. B.C. also acknowledges the generous support of the Erling Persson Family Foundation.

Funding

Open access funding provided by Stockholm University.

Author information

Authors and Affiliations

1. Stockholm Resilience Centre, Stockholm University, Stockholm, Sweden

   Beatrice I. Crona, Emmy Wassénius, Malin Jonell, Rebecca Short & Tim M. Daw

2. Global Economic Dynamics and the Biosphere, Royal Swedish Academy of Science, Stockholm, Sweden

   Beatrice I. Crona, Emmy Wassénius, Malin Jonell & Max Troell

3. Stanford Center for Ocean Solutions, Stanford University, Stanford, CA, USA

   J. Zachary Koehn, Michelle Tigchelaar, James P. Leape, Fiorenza Micheli, Elizabeth R. Selig & Colette C. C. Wabnitz

4. Dept. of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

   Christopher D. Golden

5. Dept. of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Christopher D. Golden

6. Dept. of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Christopher D. Golden

7. Dept. of Environmental Science, American University, Washington, DC, USA

Jessica A. Gephart

8. WorldFish, Bayan Lepas, Malaysia

Edward H. Allison, Michael Phillips & Shakuntala H. Thilsted

9. Wageningen University and Research, Wageningen, The Netherlands

Simon R. Bush

10. School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

Ling Cao

11. Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, British Columbia, Canada

William W. L. Cheung, U. Rashid Sumaila & Colette C. C. Wabnitz

12. EAT, Oslo, Norway

Fabrice DeClerck

13. Bloomberg School of Public Health, Berman Institute of Bioethics, Johns Hopkins University, Washington DC, USA

Jessica Fanzo

14. Nitze School of Advanced International Studies, Johns Hopkins University, Washington, DC, USA

Jessica Fanzo

15. Instituto Milenio en Socio-Ecologia Costera, Pontificia Universidad Católica de Chile, Santiago, Chile

Stefan Gelcich

16. Center of Applied Ecology and Sustainability, Pontificia Universidad Católica de Chile, Santiago, Chile

Stefan Gelcich

17. International Food Policy Research Institute (IFPRI), New Delhi, India

Avinash Kishore

18. National Center for Ecological Analysis and Synthesis, UC Santa Barbara, Santa Barbara, CA, USA

Benjamin S. Halpern

19. Bren School of Environmental Science and Management, UC Santa Barbara, Santa Barbara, CA, USA

Benjamin S. Halpern

20. Lancaster Environment Centre, Lancaster University, Lancaster, UK

Christina C. Hicks

21. Institute of Aquaculture, University of Stirling, Stirling, UK

David C. Little

22. Hopkins Marine Station, Oceans Department, Stanford University, Pacific Grove, CA, USA

Fiorenza Micheli

23. Department of Earth System Science, Stanford University, Stanford, CA, USA

Rosamond L. Naylor

24. Center on Food Security and the Environment, Stanford University, Stanford, CA, USA

Rosamond L. Naylor

25. Oxford Martin Programme on the Future of Food, University of Oxford, Oxford, UK

Marco Springmann

26. Nuffield Department of Population Health, University of Oxford, Oxford, UK

Marco Springmann

27. School of Public Policy and Global Affairs, The University of British Columbia, Vancouver, British Columbia, Canada

U. Rashid Sumaila

28. Beijer Institute of Ecological Economics, Royal Swedish Academy of Science, Stockholm, Sweden

Max Troell

Contributions

B.C. conceptualized the study, with substantial methodological and design input from T.D., E.W., M. Tigchelaar, M.J., R.S. and J.Z.K. as well as

C.D.G. and J.A.G. Data acquisition and compilation was conducted by E.W., M. Tigchelaar, M.J., R.S. and J.Z.K., and analysis was conducted by B.C. and E.W., with input from T.D. B.C. drafted the original manuscript and all co-authors reviewed and revised the writing and interpretation of findings.

Corresponding author

Correspondence to Beatrice I. Crona.

Ethics declarations

Competing interests

R.S. sits on the board of Oceana, and C.D.G., R.L.N. and J.A.G. serve as scientific advisers to the same organization. S.R.B. has unpaid advisory roles on the International Advisory Board for Aquaculture Investments of the Sustainable Trade Initiative (IDH), Utrecht, The Netherlands; and on the Standards Oversight Committee of the Global Seafood Alliance, United States; and is part of the Seafood Watch Aquaculture Multi Stakeholder Group at Monterey Bay Aquarium, CA, USA. B.C., M. Troell, E.R.S. and C.C.C.W. provide occasional voluntary and unpaid scientific support to the Seafood Business for Ocean Stewardship initiative (https://seabos.org/). None of the non-academic actors mentioned has had any input to the study design, analysis, interpretation of data or conclusions drawn.

Peer review

Peer review information

Nature thanks Daniele Brigolin, Ramón Filgueira, Andrea Hicks, Nanna Roos and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Underlying distribution of variables.

Red line indicates selected cut-offs used in our analysis. In cases where many countries have data close to the cutoff – a change in threshold value will greatly impact the outcomes, thus explaining some of the results of the sensitivity analyses (see Extended Data Figs. 2–6).

Extended Data Fig. 2 Sensitivity analysis of policy "Reducing blue food sensitive nutrient deficiencies" – for vitamin B12.

Shows number of countries in each category of policy relevance (highly relevant, relevant, less relevant), under all possible values of the threshold. Blue food variable is shown both in its full extent and in a cropped version to highlight the variability around the selected threshold. Red vertical line indicates selected threshold in analysis.

Extended Data Fig. 3 Sensitivity analysis of policy "Reducing blue food sensitive nutrient deficiencies" – for omega-3.

Shows number of countries in each category of policy relevance (highly relevant, relevant, less relevant), under all possible values of the threshold. Blue food variable is shown both in its full extent and in a cropped version to highlight the variability around the selected threshold. Red vertical line indicates selected threshold in analysis.

Extended Data Fig. 4 Sensitivity analysis of policy "Reducing cardiovascular disease risk".

Shows number of countries in each category of policy relevance (highly relevant, relevant, less relevant), under all possible values of the threshold. Blue food variable is shown both in its full extent and in a cropped version

to highlight the variability around the selected threshold. Red vertical line indicates selected threshold in analysis.

Extended Data Fig. 5 Sensitivity analysis of policy "Reducing environmental footprints of food consumption and production".

Shows number of countries in each category of policy relevance (highly relevant, relevant, less relevant), under all possible values of the threshold. Blue food variable is shown both in its full extent and in a cropped version to highlight the variability around the selected threshold. Red vertical line indicates selected threshold in analysis.

Extended Data Fig. 6 Sensitivity analysis of policy "Safeguarding food system contributions under climate change".

Shows number of countries in each category of policy relevance (highly relevant, relevant, less relevant), under all possible values of the threshold. Red vertical line indicates selected threshold in analysis.

Supplementary information

Supplementary Information

This file contains Supplementary Tables 1 and 2 and References.

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Crona, B.I., Wassénius, E., Jonell, M. *et al.* Four ways blue foods can help achieve food system ambitions across nations. *Nature* **616**, 104–112 (2023). https://doi.org/10.1038/s41586-023-05737-x

- Received: 24 February 2021
- Accepted: 17 January 2023
- Published: 22 February 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05737-x

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

From science to policy objectives

- Annisa Chand

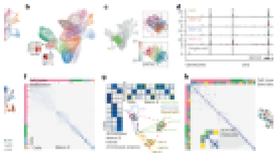
Nature Food Research Highlight 24 Mar 2023

Blue foods brought to the table to improve fish-policy decisions

- Nanna Roos

Nature News & Views 27 Mar 2023

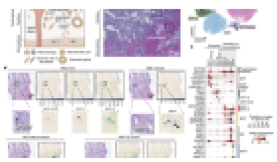
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05737-x



SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data

27 March 2023

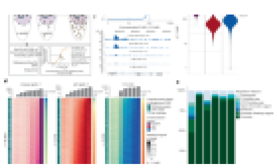
Sitara Persad, Zi-Ning Choo, … Dana Pe’er



Spatial multiomics map of trophoblast development in early pregnancy

29 March 2023

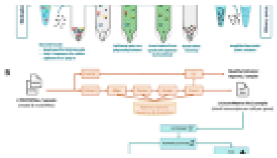
Anna Arutyunyan, Kenny Roberts, … Roser Vento-Tormo



Quantification of absolute transcription factor binding affinities in the native chromatin context using BANC-seq

27 March 2023

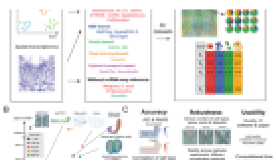
Hannah K. Neikes, Katarzyna W. Kliza, … Michiel Vermeulen



[RevGel-seq: instrument-free single-cell RNA sequencing using a reversible hydrogel for cell-specific barcoding]

24 March 2023

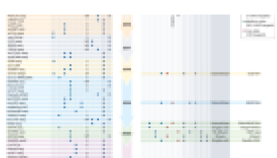
Jun Komatsu, Alba Cico, … Stuart Edelstein



A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics

21 March 2023

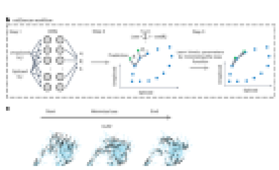
Haoyang Li, Juexiao Zhou, … Xin Gao



Methods and applications for single-cell and spatial multi-omics

02 March 2023

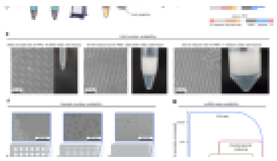
Katy Vandereyken, Alejandro Sifrim, … Thierry Voet



[A relay velocity model infers cell-dependent RNA velocity]

03 April 2023

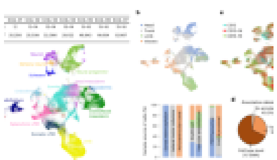
Shengyu Li, Pengzhi Zhang, … Guangyu Wang



[Microfluidics-free single-cell genomics with templated emulsification]

06 March 2023

Iain C. Clark, Kristina M. Fontanez, … Adam R. Abate



[A single-cell transcriptome atlas profiles early organogenesis in human embryos]

16 March 2023

Yichi Xu, Tengjiao Zhang, … Weiyang Shi

- Article
- Open Access
- [Published: 15 March 2023]

Spatial epigenome–transcriptome co-profiling of mammalian tissues

- Di Zhang,
- Yanxiang Deng ,
- Petra Kukanja,
- Eneritz Agirre,
- Marek Bartosovic,
- Mingze Dong,
- Cong Ma,
- Sai Ma,
- Graham Su,
- Shuozhen Bao,
- Yang Liu,
- Yang Xiao,
- Gorazd B. Rosoklija,
- Andrew J. Dwork,
- J. John Mann,
- Kam W. Leong,
- Maura Boldrini,
- Liya Wang,
- Maximilian Haeussler,
- Benjamin J. Raphael,
- Yuval Kluger,
- Gonçalo Castelo-Branco &
- …
- Rong Fan

Show authors

Nature **volume 616**, pages 113–122 (2023)

- 43k Accesses
- 2 Citations
- 280 Altmetric
- Metrics details

Abstract

Emerging spatial technologies, including spatial transcriptomics and spatial epigenomics, are becoming powerful tools for profiling of cellular states in the tissue context[1,2,3,4,5]. However, current methods capture only one layer of omics information at a time, precluding the possibility of examining the mechanistic relationship across the central dogma of molecular biology. Here, we present two technologies for spatially resolved, genome-wide, joint profiling of the epigenome and transcriptome by cosequencing chromatin accessibility and gene expression, or histone modifications (H3K27me3, H3K27ac or H3K4me3) and gene expression on the same tissue section at near-single-cell resolution. These were applied to embryonic and juvenile mouse brain, as well as adult human brain, to map how epigenetic mechanisms control transcriptional phenotype and cell dynamics in tissue. Although highly concordant tissue features were identified by either spatial epigenome or spatial transcriptome we also observed distinct patterns, suggesting their differential roles in defining cell states. Linking epigenome to transcriptome pixel by pixel allows the uncovering of new insights in spatial epigenetic priming, differentiation and gene regulation within the tissue architecture. These technologies are of great interest in life science and biomedical research.

Main

Single-cell multiomics allows uncovering of the mechanisms of gene regulation across different omics layers[6,7,8,9] but lacks spatial information, which is crucial to understanding cellular function in tissue. Spatial epigenomics, transcriptomics and proteomics emerged recently[1,2,3,4,5] but most of these can capture only one layer of the omics information. Although computational methods can integrate data from multiple omics[10], they cannot readily uncover the mechanistic links between different omics layers. Previously we developed deterministic barcoding in tissue for spatial omics sequencing for spatially resolved comeasurement of the transcriptome and a panel of proteins[1], and recently it was also realized in the 10X Visium platform[11]. Herein, to further investigate the epigenetic mechanisms underlying gene expression regulation, we developed spatially resolved, genome-wide comapping of the epigenome and transcriptome by simultaneous profiling of chromatin accessibility and messenger RNA

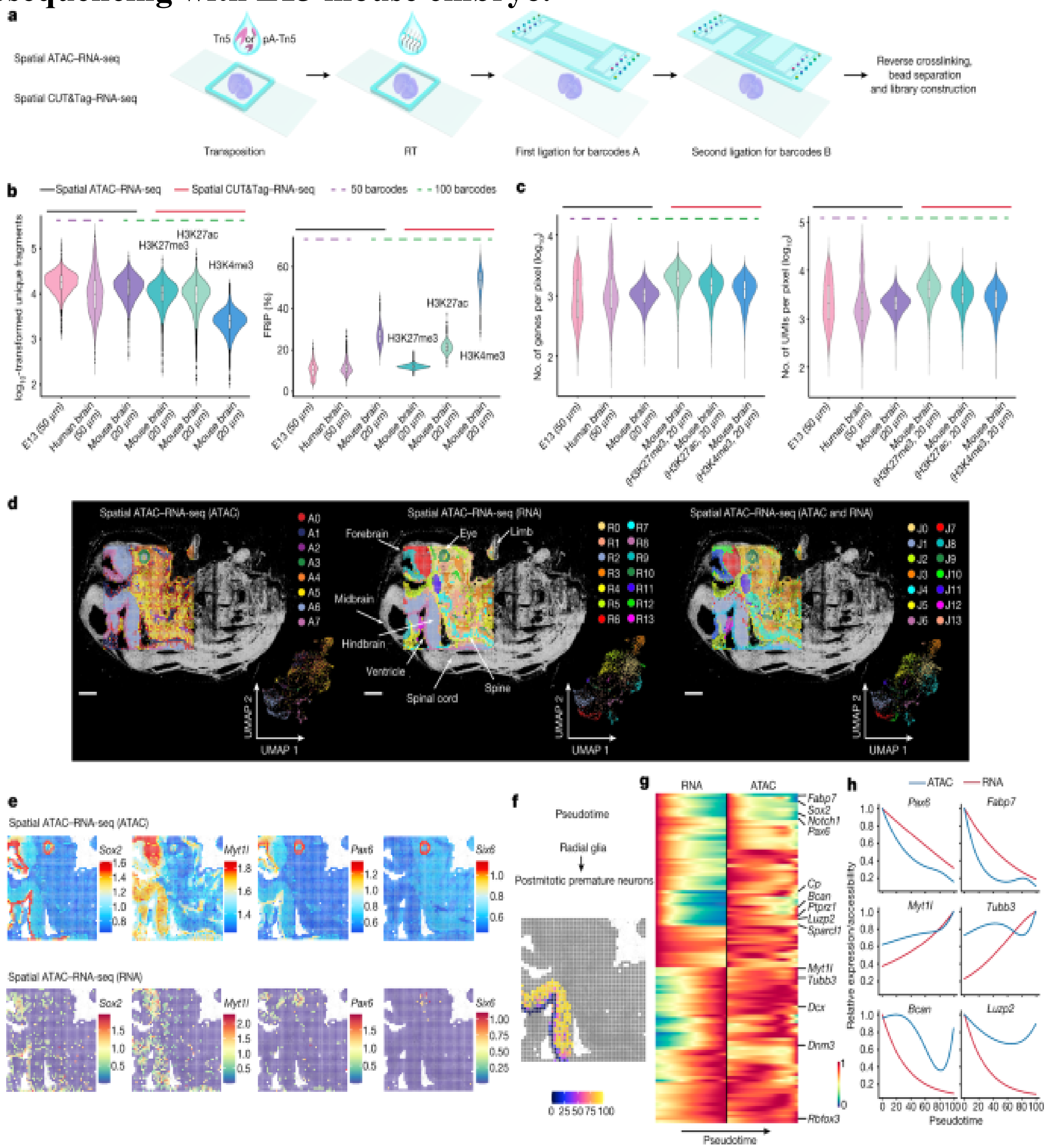
expression (spatial assay for transposase-accessible chromatin and RNA using sequencing (spatial ATAC–RNA-seq)), or histone modifications and mRNA expression (spatial assay of cleavage under targets and tagmentation and RNA using sequencing (spatial CUT&Tag–RNA-seq); applied to H3K27me3, H3K27ac or H3K4me3 histone modifications) on the same tissue section at the cellular level via deterministic cobarcoding, to integrate the chemistry of spatial-ATAC-seq[3] or spatial-CUT&Tag[2] with that for spatial transcriptomics. We applied these techniques to embryonic and juvenile mouse brain, as well as adult human brain hippocampus, to dissect epigenetic and transcriptional states and their role in the regulation of cell types dynamics in tissue. This work opens new frontiers in spatial omics and may provide unprecedented opportunities to biological and biomedical research.

Technology workflow and data quality

Spatial ATAC–RNA-seq is shown schematically in Fig. 1a and Extended Data Fig. 1a–c. A frozen tissue section was fixed with formaldehyde and treated with Tn5 transposition complex preloaded with a DNA adaptor containing a universal ligation linker that can be inserted into transposase-accessible genomic DNA loci. The same tissue section was then incubated with a biotinylated DNA adaptor containing a universal ligation linker and a poly-T sequence that binds to the poly-A trail of mRNAs to initiate reverse transcription (RT) in tissue. A microfluidic channel array chip was then placed on the tissue section to introduce spatial barcodes Ai (i = 1−50 or 100) that were covalently conjugated to the universal ligation linker via templated ligation. Next, another microchip with microchannels perpendicular to the first flow direction was used to introduce spatial barcodes Bj (j = 1−50 or 100) that were then ligated to barcodes Ai (i = 1−50 or 100), resulting in a two-dimensional grid of spatially barcoded tissue pixels, each defined by the unique combination of barcodes Ai and Bj (i = 1−50 or 100, j = 1−50 or 100; barcoded pixels, n = 2,500 or 10,000). Finally, barcoded complementary DNA and genomic DNA fragments were released after reverse crosslinking. cDNAs were enriched with streptavidin beads and gDNA fragments were retained in the supernatant. The libraries of gDNA and cDNA were constructed separately for next-generation sequencing (NGS). Spatial CUT&Tag–RNA-seq was performed by applying an

antibody against a specific histone modification (H3K27me3, H3K27ac or H3K4me3) to the tissue section and then using a protein A-tethered Tn5-DNA complex to perform CUT&Tag. The remaining steps were similar to those for spatial ATAC–RNA-seq (Fig. 1a and Extended Data Fig. 1a,d,e), but resulted in spatial co-profiling of genome-wide histone modification occupancy and transcriptome.

Fig. 1: Design and evaluation of spatial epigenome–transcriptome cosequencing with E13 mouse embryo.



a, Schematic workflow. **b**, Comparison of number of unique fragments and fraction of reads in peaks (FRiP) in spatial ATAC–RNA-seq and spatial CUT&Tag–RNA-seq. **c**, Gene and UMI count distribution in spatial ATAC–RNA-seq and spatial CUT&Tag–RNA-seq. Number of pixels in E13, 2,187; in human brain, 2,500; in mouse brain (ATAC), 9,215; in mouse brain (H3K27me3), 9,752; in mouse brain (H3K27ac), 9,370; in mouse brain (H3K4me3), 9,548. Box plots show the median (centre line), the first and third quartiles (box limits) and 1.5× interquartile range (whiskers). **d**, Spatial distribution and UMAP of all clusters for ATAC, RNA and joint clustering of ATAC and RNA data. Overlay of clusters with the tissue image shows that spatial clusters precisely match anatomic regions. Pixel size, 50 µm; scale bars, 1 mm. **e**, Spatial mapping of GAS and gene expression for selected marker genes in different clusters for ATAC and RNA in spatial ATAC–RNA-seq. **f**, Pseudotime analysis from radial glia to postmitotic premature neurons visualized at the spatial level. **g**, Heatmaps delineating gene expression and GAS for marker genes. **h**, Dynamic changes in GAS and gene expression across pseudotime.

We performed spatial ATAC–RNA-seq experiments on embryonic day 13 (E13) mouse embryo (pixel size, 50 µm), mouse postnatal day 21/22 (P21/22) brains (pixel size, 20 µm) and adult human brain hippocampus tissue (pixel size, 50 µm). For the 50-µm pixel size we obtained an average of 25 cells for E10 mouse embryo[1] and between one and nine cells for human hippocampus[3]. Most 20-µm pixels contained one to three cells per pixel in juvenile mouse brain (Supplementary Fig. 1b). Using a 100 × 100 barcode scheme, the mapping area covers nearly the entire hemisphere of a P22 mouse brain coronal section. From this sample we obtained a median of 14,284 unique fragments per pixel, of which 19% were enriched in the transcription start site regions and 26% located in peaks (Fig. 1b and Supplementary Figs. 1a and 2a). For the RNA portion a total of 22,914 genes were detected, with an average of 1,073 genes and 2,358 unique molecular identifiers (UMIs) per pixel (spatial ATAC–RNA-seq) (Fig. 1c). We performed spatial CUT&Tag–RNA-seq for H3K27me3, H3K27ac and H3K4me3 on mouse P21/22 brains (pixel size, 20 µm). With the 100 × 100 barcode device we obtained a median of 10,644 (H3K27me3), 10,002 (H3K27ac) and 2,507 (H3K4me3) unique fragments per pixel, of which 12% (H3K27me3), 17% (H3K27ac) and 67% (H3K4me3) overlapped with

transcription start site regions and 12% (H3K27me3), 21% (H3K27ac) and 54% (H3K4me3) were located in peaks, respectively (Fig. 1b and Supplementary Figs. 1a and 2a). For the RNA data, 25,881 (H3K27me3), 23,415 (H3K27ac) and 22,731 (H3K4me3) genes were detected with an average of 2,011 (H3K27me3), 1,513 (H3K27ac) and 1,329 (H3K4me3) genes per pixel (4,734 (H3K27me3), 3,580 (H3K27ac) and 2,885 (H3K4me3) UMIs per pixel, respectively (Fig. 1c)). Assessment of data quality for mouse embryo, P21 mouse brain and human brain samples with 50 × 50 barcodes is also included in Fig. 1b,c, Supplementary Figs. 1a and 2a–c and Supplementary Tables 2 and 3.

Moreover, the insert size distributions of chromatin accessibility (spatial ATAC–RNA-seq) and histone modification (spatial CUT&Tag–RNA-seq, H3K27me3, H3K27ac or H3K4me3) fragments were consistent with the captured nucleosomal fragments in all tissues (Supplementary Fig. 1c,d). Correlation analysis between replicates showed high reproducibility ($r = 0.98$ for ATAC, $r = 0.98$ for RNA in spatial ATAC–RNA-seq, $r = 0.96$ for CUT&Tag (H3K27ac), $r = 0.89$ for RNA in spatial CUT&Tag–RNA-seq; r denotes Pearson correlation coefficient; Supplementary Fig. 2d,e). Tissue type, preparation and quality can all influence analytical metrics (Methods).

Spatial comapping of mouse embryo

Spatial ATAC–RNA-seq on E13 mouse embryo identified eight major ATAC clusters and 14 RNA clusters, suggesting that, at this stage of development, chromatin accessibility may not allow discrimination of all cell types defined by transcriptional profiles. Spatial distribution of these clusters agrees with tissue histology (see haematoxylin and eosin staining of an adjacent tissue section in Fig. 1d and Extended Data Fig. 2a). In the ATAC data, cluster A3 represents the embryonic eye field with open chromatin accessibility at the loci of genes including *Six6* (Fig. 1d (left) and Fig. 1e). Clusters A4–A5 are associated with several developing internal organs. Clusters A6–A7 cover the central nervous system (CNS). To benchmark the ATAC result, we aggregated the chromatin accessibility profiles in pixels within specific organs and compared these with organ-specific ENCODE E13.5 ATAC-seq reference data (Supplementary Fig. 3b). Additionally, the peaks obtained from our spatial ATAC data are also consistent with the

ENCODE reference (Supplementary Fig. 3c). We integrated spatial ATAC data with single-cell RNA-seq[12] data for cell type assignment in each cluster (Extended Data Fig. 2c,d). As expected, radial glia (neural stem/progenitor cells) were observed predominantly in the ventricular layer, and differentiated cell types such as postmitotic premature neurons and inhibitory interneurons were enriched in those regions distant from the ventricular layer (Extended Data Fig. 2d).

Cell type-specific marker genes were identified for individual clusters, and their expression was inferred from chromatin accessibility (Fig. 1e and Extended Data Fig. 2b,e) and predicted by gene activity score (GAS[13]; Methods). *Sox2*, which is involved in the development of nervous tissue and optic nerve formation[14,15], showed high chromatin accessibility in the embryonic eye field and in the ventricular layer containing neural stem/progenitor cells. *Pax6* exhibited a similar spatial pattern of chromatin accessibility. *Myt1l*, which encodes myelin transcription factor 1-like protein, presented a higher ATAC signal in the embryonic brain and neural tube[16]. *Six6*, a key gene involved in eye development[17], showed highest GAS in the eye region. *Nrxn2*, which encodes Neurexin 2, a key gene in the vertebrate nervous system[18], was extensively accessible in most neural cell regions. Accessibility at *Rbfox3*, which encodes RNA binding protein fox-1 homolog 3 (a splicing factor known as NeuN)[19], was observed in neurons[19] whereas *Sox1* and *Sox2* presented enriched chromatin accessibility in the ventricular layer (Fig. 1e and Extended Data Fig. 2b). Cell type-specific enrichment of transcription factor (TF) regulators was also examined using ChromVAR[20] analysis of deviation in TF motifs (*Sox2* and *Nfix*), and identified the positive TF regulators (Extended Data Fig. 2g,h). We observed that the *Sox2* motif was enriched in cluster A7, consistent with its function in embryonic brain development[21]. GREAT analysis[22] further verified the strong concordance between gene regulatory pathways and anatomical annotation (Extended Data Fig. 2i,j).

For the RNA spatial data, 14 distinct clusters were identified and characterized by specific marker genes (Fig. 1d (middle) and Extended Data Fig. 2f). For example, cluster R10 (*Six6*) was correlated with the embryonic eye, clusters R2, R6 and R8 were related to the CNS and cluster R7 was associated with the formation of cartilage. To evaluate data quality we

performed correlation analysis with organ-specific ENCODE E13.5 RNA-seq reference data, which showed high concordance (Supplementary Fig. 3a). Cell type-specific marker genes were also identified for individual RNA clusters—for example, *Mapt* (R2) may function in establishing and maintaining neuronal polarity[23]. *Epha5* (R6) is involved in axon guidance, whereas *Slc1a3* (R8) is involved in regulation of excitatory neurotransmission in the CNS[18]. *Myh3* (R3) is associated with muscle contraction[24]. *Col2a1* (R7), which shows specific expression in cartilaginous tissues, has an essential role in normal embryonic skeleton development[18]. Pathway analysis[25] results agree with anatomical annotation (Supplementary Fig. 4a).

Integration of spatial RNA data with scRNA-seq mouse organogenesis data[12] was performed to determine cell identities in each pixel (Extended Data Fig. 2c,d). We observed that radial glia, postmitotic premature neurons and inhibitory interneurons (clusters R8, R6 and R2) were present in the same major clusters as shown in the ATAC analysis (clusters A7 and A6), which verified the use of multiple omics information for more robust cell type identification. We attempted joint clustering of spatial ATAC and RNA data to refine the spatial patterns. A new neuronal cluster (J10) was identified in the joint clustering analysis, which was not readily resolved by single modalities alone (Fig. 1d, right). This result highlights the value of using joint multiomics profiles to improve spatial cell type mapping[26].

Co-profiling of the epigenome and transcriptome facilitates investigation of the correlation between accessible peaks and expressed genes pixel by pixel in the tissue context. We observed distinct signals at predicted enhancers for some genes (*Sox2*, *Pax6* and *Sox1*) (Extended Data Fig. 2e and Supplementary Fig. 4b). For example, enhancers for *Sox2* and *Pax6* had higher chromatin accessibility in clusters A7 and A3, respectively, suggesting their roles in these tissue regions in the regulation of *Sox2* and *Pax6* expression. Although spatial RNA distribution of these genes corresponded with their chromatin accessibility, the expression level may differ significantly from the degree of accessibility (Fig. 1e and Extended Data Fig. 2b). Some marker genes (*Pax6*, *Sox2* and *Myt1l*) were highly accessible in some regions of the embryonic brain but showed modest or low RNA expression (Fig. 1e and Extended Data Fig. 2b), which may indicate

lineage priming[10,27] of these genes in embryonic brain development. Despite a strong correlation between replicates (Supplementary Fig. 2d), this observation could be due in part to the sequencing depth and RNA detection sensitivity. Nevertheless, these results still highlight the potential to link genome-wide epigenetic regulation to transcription in the spatial tissue context.

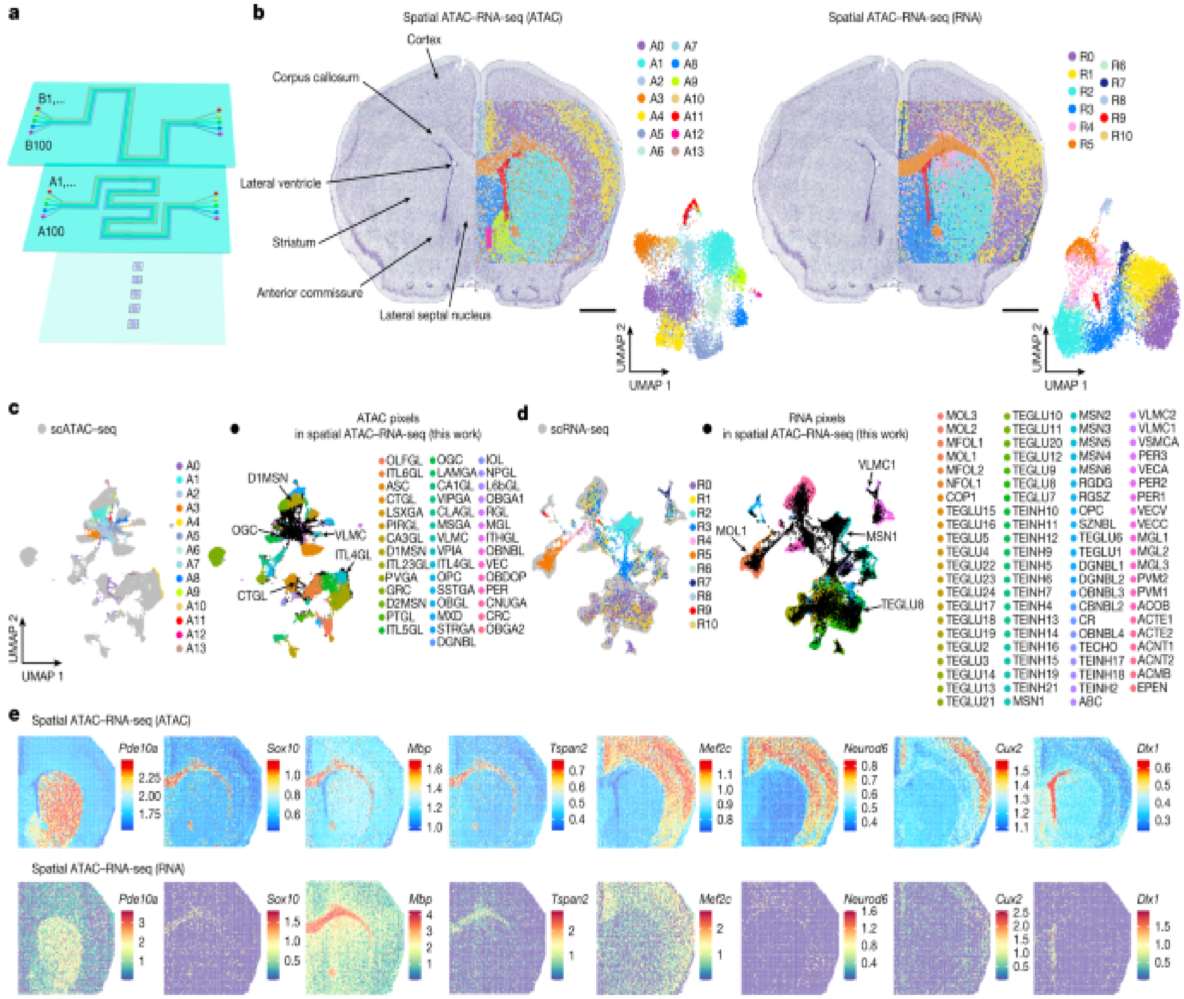
To investigate the spatiotemporal relationship between chromatin accessibility and gene expression in embryonic development, we analysed the differentiation trajectory from radial glia to various types of neurons such as postmitotic premature neurons[28]. Pseudotime analysis[29] was conducted under the ATAC pseudotime coordinate system, and developmental trajectories were directly visualized in the spatial tissue map (Fig. 1f). Chromatin accessibility GAS and gene expression along this trajectory showed dynamic changes in selected marker genes (Fig. 1g,h). As expected, the expression levels of *Sox2*, *Pax6* and other genes involved in progenitor maintenance and proliferation (Gene Ontology (GO) *Fabp7* to *Pax6* in Extended Data Fig. 3a) were downregulated during the transition to postmitotic neurons. The loss of chromatin accessibility at the *Pax6* and radial glia marker *Fabp7* loci preceded downregulation of the corresponding RNAs (Fig. 1h). In turn, genes involved in neuronal identity, axonogenesis and synapse organization (GO *Myt1l* to *Dnm3* in Extended Data Fig. 3a), such as *Dcx* and *Tubb3*, showed increased expression in spatial pseudotime but chromatin accessibility at their loci was already elevated at earlier stages, suggesting lineage priming of these genes in expression[10,27]. We also found a cohort of genes whose expression rapidly declined during spatial pseudotime but whose chromatin accessibility was maintained throughout pseudotime or declined only at very late stages. Many of these genes, such as *Ptprz1*, *Bcan* and *Luzp2*, are characteristic of oligodendrocyte precursor cells, and the corresponding GO biological processes are negative regulation of myelination and regulation of gliogenesis (GO *Cp* to *Sparcl1* in Extended Data Fig. 3a), suggesting that the neuronal lineage might retain the potential to acquire an oligodendrocyte identity even when it has already migrated away from the ventricular zone in the embryonic brain. Monocle2 pseudotime analysis showed a bifurcation in chromatin accessibility but not in RNA expression; one path led to regions close to the ventricular zone (green pixels, Extended Data Fig. 3d,e) and the other terminated in regions

distal to the ventricular zone (blue pixels). In contrast to the green path, the blue path presented an increase in chromatin accessibility for genes involved in axonogenesis and dendrite formation (Extended Data Fig. 3e (red box), 3f), suggesting that the chromatin state of neural cells distal to the ventricle corresponds to a more differentiated neuronal state. Thus, our spatial ATAC–RNA-seq can be used to decipher the gene regulation mechanism and spatiotemporal dynamics during tissue development.

Spatial ATAC–RNA-seq of mouse brain

We developed a microchip with 100 serpentine microchannels to barcode 100 × 100 or, in total, 10,000 pixels per tissue section and up to five samples per chip (Fig. 2a, 20-μm pixel size). This was applied to P22 mouse brain coronal sections (at bregma 1) for joint profiling of chromatin accessibility with transcriptome. In contrast to E13 mouse embryonic brain, we found a higher number of ATAC clusters (14) compared with RNA clusters (11), indicating terminal differentiation of most major cell types in the juvenile brain relative to that in embryo, which consists of many undifferentiated or multipotent cell states. Spatial distribution of clusters agreed with the anatomical annotation defined by Nissl staining, reflecting arealization of the juvenile brain (Fig. 2b). Moreover, spatial clusters between ATAC and RNA showed concordance in cluster assignment (Supplementary Fig. 5a). Chromatin accessibility of marker genes identified major regions such as striatum (*Pde10a* and *Adcy5*, markers of medium spiny neurons, cluster A1), corpus callosum (*Sox10*, *Mbp* and *Tspan2*, markers of oligodendroglia, cluster A3), cortex (*Mef2c*, *Neurod6* and *Nrn1*, clusters A0 and A4, markers of excitatory neurons; *Cux2*, cluster A4) and lateral ventricle (*Dlx1*, *Pax6, Notch1* and *Sox2*, markers of ependymal/neural progenitor cells, cluster A11) (Fig. 2e and Extended Data Fig. 4a,b). GREAT analysis confirmed that the major pathways correlated with tissue functions in different anatomical regions (Supplementary Fig. 5b,c). Spatial RNA clusters also showed region-specific gene expression such as striatum (*Pde10a*, cluster R2, medium spiny neurons), corpus callosum (*Mbp* and *Tspan2*, cluster R5, oligodendroglia) and cortex (*Mef2c*, clusters R0 and R1, excitatory neurons) (Fig. 2e and Extended Data Fig. 4a).

Fig. 2: Spatial chromatin accessibility and transcriptome co-profiling of P22 mouse brain.



a, Design of microfluidic chips for 100 × 100 barcodes with 20-µm channel size. **b**, Spatial distribution and UMAP of all clusters for ATAC and RNA in spatial ATAC–RNA-seq of mouse brain. Pixel size, 20 µm; scale bars, 1 mm. **c**, Integration of ATAC data and scATAC-seq data[30] from mouse brain. **d**, Integration of RNA data and scRNA-seq data[32] from mouse brain. **e**, Spatial mapping of GAS and gene expression for selected marker genes in different clusters for ATAC and RNA in spatial ATAC–RNA-seq. A list of abbreviation definitions can be found in Supplementary Table 1.

Integration of single-cell ATAC-seq mouse brain atlas data[30] with spatial ATAC-seq data facilitated identification of all major cell types, and label transfer was then used to assign cell types to spatial locations where the epigenetic state may control specific cell type formation (Fig. 2c and

Extended Data Fig. 4c). For example, we observed immature oligodendrocytes and oligodendrocytes in cluster A3 corresponding to corpus callosum. A thin layer of radial glia-like cells was found in the lateral ventricle, medium spiny neurons (D1MSN and D2MSN) in striatum and inhibitory neurons in the lateral septal nucleus[30] (Extended Data Fig. 4c) based on chromatin accessibility. Furthermore, the subclasses of excitatory neurons in different layers of cortex (ITL23GL, Cortex L2/3; ITL4GL, Cortex L4; ITL5GL, Cortex L5; PTGL, Cortex PT; NPGL, Cortex NP; ITL6GL, Cortex L6; CTGL, Cortex CT; and L6bGL, Cortex L6b) were clearly identified[31] (Extended Data Fig. 4c). Integration of spatial RNA data and the juvenile CNS scRNA-seq atlas[32] allowed for visualization of dominant transcriptional cell types in tissue (Fig. 2d and Extended Data Fig. 4d). We observed a high degree of concordance in spatial cell type distribution as determined by ATAC versus RNA, suggesting a general congruence between chromatin accessibility and transcriptome in defining cell identities in tissue. For instance, neuronal intermediate progenitor cells (SZNBL, included in radial glial-like cells from scATAC-seq[30]), mature oligodendrocytes (MOL, included in oligodendrocytes from scATAC-seq), newly formed oligodendrocytes (included in immature oligodendrocytes from scATAC-seq) and medium spiny neurons (MSN) shown by the RNA data were enriched in the same regions as identified by the ATAC data (Extended Data Fig. 4c,d). Furthermore, the detection of a thin layer of SZNBL in the lateral ventricle and vascular leptomeningeal cells (VLMC) in the regions with preserved meninges demonstrated a high spatial resolution for our technology in identifying low-abundance cells, even at near-single-cell resolution (for VLMC) (Extended Data Fig. 4c,d).

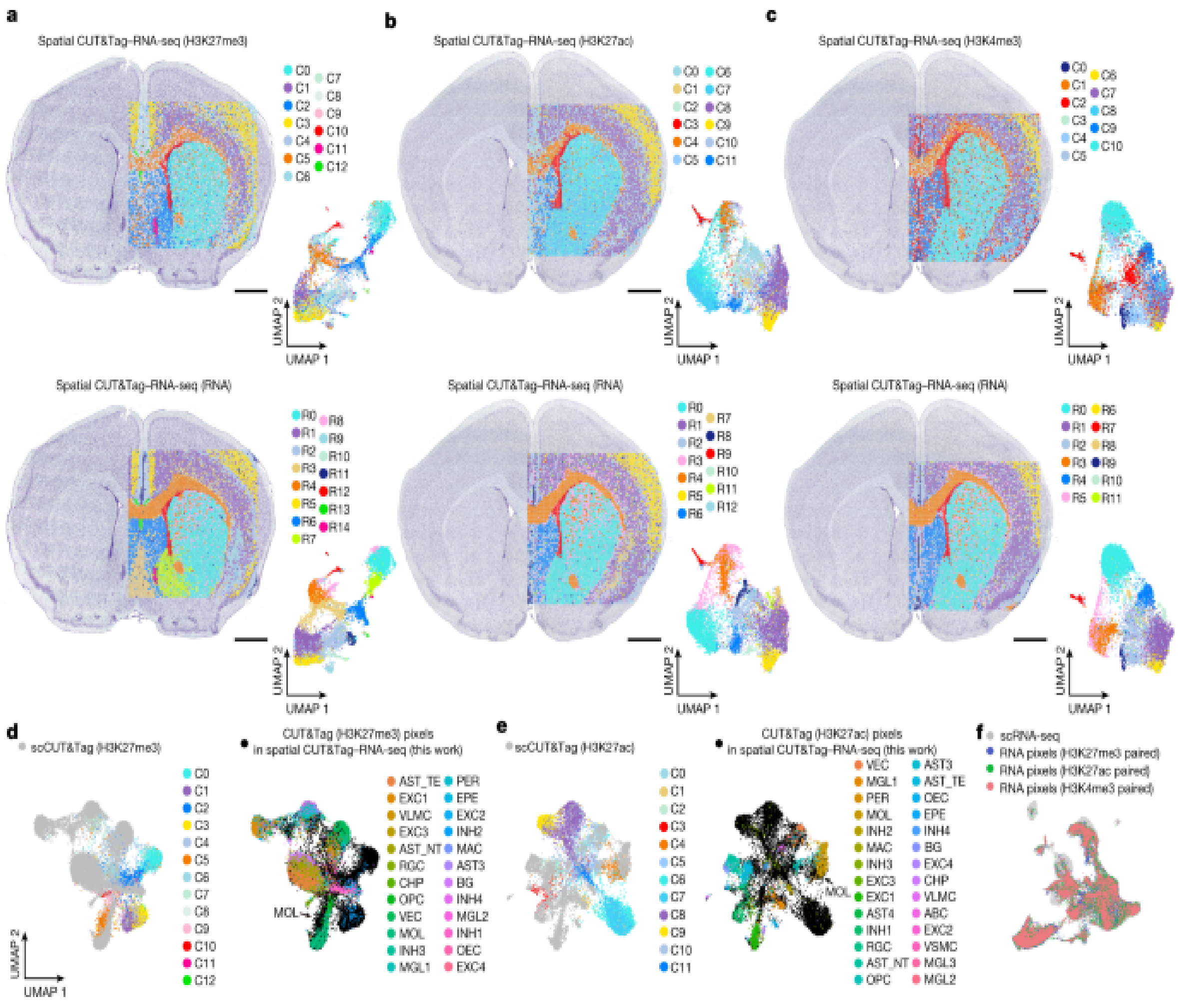
Joint profiling of ATAC and RNA facilitates inferring the gene regulatory landscape by searching correlated peak accessibility and gene expression. We detected 21,417 significant peak-to-gene linkages between regulatory elements and target genes (Extended Data Fig. 5c,d). Some potential enhancers with dynamically regulated promoter interactions were found to be enriched in specific clusters such as *Dlx1* and *Sox2* (cluster A11)*, Tspan2* (cluster A3) and *Adcy5* (cluster A1) (Extended Data Fig. 4b), indicating the ability of spatial ATAC–RNA-seq to identify the key regulatory regions for target genes. Spatial patterns of TF motif enrichment (*Dnajc21*, *Pax6*, *Sox2*, *Sox4* and *Mef2c*) provided further insights into the TF regulators[33]

visualized in tissue (Extended Data Fig. 5a,b). For example, *Sox2* was examined simultaneously for chromatin accessibility, gene expression, putative enhancers and TF motif enrichment, enabling a more comprehensive understanding of gene regulation dynamics. As observed in the embryonic mouse brain, some marker genes with open chromatin accessibility (*Sox10*, *Sox2*, *Neurod6*, *Pax6* and *Notch1*) were either not or lowly expressed (Fig. 2e and Extended Data Fig. 4a). A biological replicate experiment on P21 mouse brain was also performed (Extended Data Fig. 6; 20 μm pixel size, 50 × 50 barcodes). Many of these genes encode for transcription factors involved in neural development, suggesting the possibility of epigenetic—but not transcriptional—memory of these genes in brain development.

Spatial CUT&Tag–RNA-seq of mouse brain

In addition to chromatin accessibility, histone modifications are also a key aspect of epigenetic regulation. Spatial CUT&Tag–RNA-seq was performed to co-profile transcriptome and H3K27me3 (repressing loci), H3K27ac (activating promoters and/or enhancers) or H3K4me3 (active promoters) histone modifications, respectively, in P22 mouse brain (20-μm pixel size, 100 × 100 barcodes). We identified 13 and 15 specific clusters for H3K27me3 and RNA, 12 and 13 clusters for H3K27ac and RNA and 11 and 12 clusters for H3K4me3 and RNA, respectively (Fig. 3a–c). These clusters agreed with the anatomical annotation by Nissl staining and showed good concordance between CUT&Tag and RNA in spatial patterns (Fig. 3a–c), which was further confirmed by Belayer[34] analysis (Supplementary Fig. 7c).

Fig. 3: Spatial histone modification and transcriptome co-profiling of P22 mouse brain.



a–**c**, Spatial distribution and UMAP of all clusters for H3K27me3 and RNA (**a**), H3K27ac and RNA (**b**) and H3K4me3 and RNA (**c**) in mouse brain. Pixel size, 20 µm; scale bars, 1 mm. **d**, Integration of H3K27me3 data with scCUT&Tag (H3K27me3) data[37] from mouse brain. **e**, Integration of H3K27ac data with scCUT&Tag (H3K27ac) data[37] from mouse brain. **f**, Integration of RNA data in spatial CUT&Tag (H3K27me3)–RNA-seq, spatial CUT&Tag (H3K27ac)–RNA-seq and spatial CUT&Tag (H3K4me3)–RNA-seq with scRNA-seq data[32] from mouse brain.

For H3K27me3, gene expression was predicted by calculation of chromatin silencing score (CSS[2,35]; Methods). High CSS indicates repressed gene expression because of the transcriptional repression function of H3K27me3. For H3K27ac and H3K4me3, active genes should correspond to high GAS[2,13] (Methods). Clustering of CSS or GAS based on different modifications resolved all major tissue regions (Fig. 3a–c and

Supplementary Fig. 6) and identified region-specific marker gene modifications. *Cux2* showed high GAS in cluster C9 (H3K27ac) and cluster C6 (H3K4me3), indicating enrichment of excitatory neurons[5]. *Cux2* was observed in cortical layers 2/3, corresponding to superficial layers in the cortex. By contrast, H3K27me3 was depleted at *Cux2* in the same region (cluster C3) (Supplementary Fig. 6). GAS of *Fezf2* (a marker gene for cortical layer 5) was high in cluster C8 (H3K27ac) and cluster C7 (H3K4me3), corresponding to the deeper layer of the cortex. *Fezf2* was depleted for H3K27me3 (cluster C1) in this layer. *Satb2* showed the highest activity in clusters C8 and C9 (H3K27ac) and C6 and C7 (H3K4me3), but the lowest CSS in clusters C1 and C3 (H3K27me3) in the cortical layer. *Tspan2* was enriched in cluster C4 (H3K27ac) and cluster C1 (H3K4me3) but depleted in cluster C5 (H3K27me3) in corpus callosum, associated with oligodendrocyte lineage development (Supplementary Fig. 6). The corresponding RNA clusters also showed concordant region-specific signatures (Fig. 3a–c and Supplementary Fig. 6), as exemplified by *Cux2* in cortical layer 2/3 excitatory neurons (cluster R5 (H3K27me3 paired), cluster R5 (H3K27ac paired) and cluster R6 (H3K4me3 paired)), *Fezf2* in cortical layer 5 excitatory neurons (cluster R1 (H3K27me3 paired), cluster R1 (H3K27ac paired), cluster R1 (H3K4me3 paired)), *Tspan2* in oligodendrocyte lineage cells in corpus callosum (cluster R4 (H3K27me3 paired), cluster R4 (H3K27ac paired) and cluster R3 (H3K4me3 paired) Supplementary Fig. 6).

Integration and co-embedding of spatial CUT&Tag and scCUT&Tag data[36,37] (Fig. 3d,e and Extended Data Fig. 7a) showed that the epigenetic states in our data for H3K27me3, H3K27ac and H3K4me3 agreed with the corresponding projection in scCUT&Tag. Integration with the corresponding juvenile CNS scRNA-seq atlas[32] allowed for label transfer to assign transcriptional cell types to the spatial location of epigenetic identities/states (Extended Data Fig. 7g,h, bottom). For example, we observed an enrichment of MOL within the corpus callosum, a thin layer of ependymal cells in the lateral ventricle, excitatory neurons in cerebral cortex and MSN in the striatum. We also integrated paired RNA data with the scRNA-seq atlas[32] to identify dominant transcriptional cell types via label transfer (Fig. 3f and Extended Data Fig. 7b–h). MOL, a thin layer of ependymal cells, MSN and excitatory neurons were enriched in the same spatial regions identified by

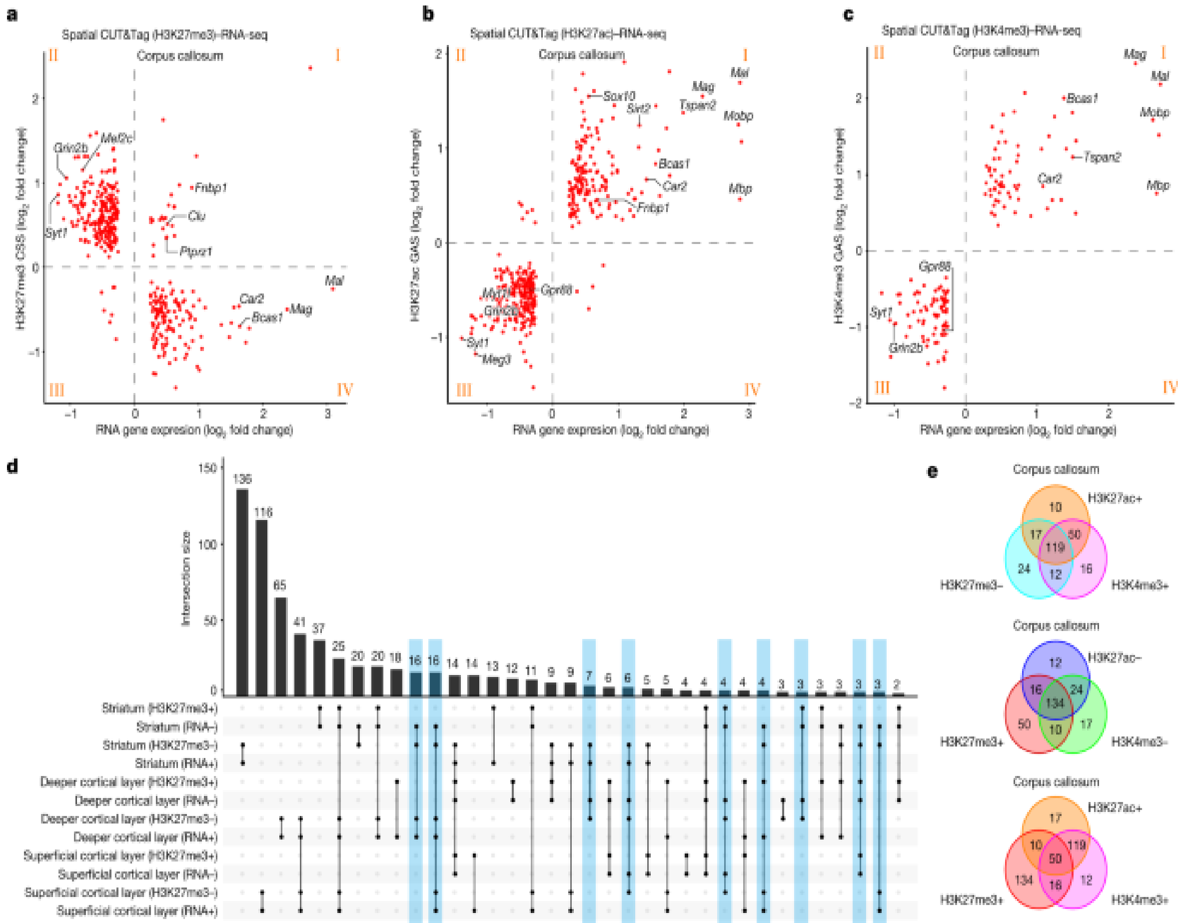
CUT&Tag (H3K27ac and H3K4me3) (Extended Data Fig. 7g,h, top). In particular for spatial H3K27me3, whereas integration with scCUT&Tag could not clearly show the identity of several clusters in the epigenomic modalities (clusters C0, C1 and C3), label transfer of scRNA-seq data with paired RNA data from the same tissue section clearly showed cell identities in these clusters (Fig. 3a,d and Extended Data Fig. 7f), highlighting the power of combining CUT&Tag and RNA-seq (spatial CUT&Tag–RNA-seq) in the same tissue section to more accurately identify cell types or states. To directly infer interactions between genome-wide gene expression and the corresponding enhancers across all clusters, we identified a total of 19,468 significant peak-to-gene linkages from spatial CUT&Tag (H3K27ac)–RNA-seq (Supplementary Fig. 7a,b). A biological replicate was also performed on P21 mouse brain (Extended Data Fig. 8; 20-µm pixel size, 50 × 50 barcodes).

Region-specific gene expression regulation

To further understand the spatial epigenetic regulation of gene expression at the genome scale, we compared the GAS or CSS obtained from ATAC and CUT&Tag (H3K27me3, H3K27ac and H3K4me3) with the corresponding RNA expression in all major tissue regions of P22 mouse brain (Fig. 4a–c and Supplementary Figs. 9a, 10a and 11a). For example, in oligodendrocyte-abundant corpus callosum we observed a robust anticorrelation between H3K27me3 and RNA (Fig. 4a). Genes such as *Mal*, *Mag* and *Car2* with low CSS and high RNA expression correspond to GO biological processes such as myelination and regulation of oligodendrocyte differentiation (Extended Data Fig. 9a–c and Supplementary Fig. 8; GO for quadrant IV). By contrast, genes such as *Grin2b* and *Syt1* with high CSS and low RNA expression are related to neuronal processes such as synaptic transmission and neurotransmitter release (Extended Data Fig. 9d,e and Supplementary Fig. 8; GO for quadrant II). These results are in accordance with ATAC/H3K27me3/H3K27ac Nano-CUT&Tag analysis[37] of oligodendroglial differentiation in the juvenile brain, which showed two waves of H3K27me3 repression during this process[37]. We also found a small subset of genes (Fig. 4a and Supplementary Fig. 8; GO for quadrant I) with higher levels of both H3K27me3 and RNA in the corpus callosum, such as *Ptprz1*, a marker of oligodendrocyte precursor cells[38,39] (Extended Data Fig. 9g), which might

indicate transcriptional poising of some of these genes. This could also be due in part to the presence in proximity of both oligodendrocyte precursor cells (expressing *Ptprz1*) and mature oligodendrocytes (where *Ptprz1* is repressed) in the corpus callosum. We performed a similar correlation analysis for RNA and H3K27me3 in other regions of the juvenile brain, including the striatum and superficial and deeper cortical layers (Supplementary Figs. 9a, 10a and 11a). We also observed strong anticorrelation with a cohort of genes involved in neuronal processes being activated or repressed in a region-specific manner. For instance, in the superficial cortical layer, genes for GABAergic regulation of synaptic transmission were enriched in H3K27me3 whereas glutamatergic synapse transmission genes had high expression and low H3K27me3 (Supplementary Figs. 9b, 10b and 11b; GOs for striatum, superficial and deeper cortical layers). In contrast to corpus callosum, only a limited number of genes positive for both H3K27me3 and RNA were found in these regions. Despite a general anticorrelation between RNA expression and H3K27me3 deposition, we also observed regional differences (blue-shaded columns in Fig. 4d); *Nav3* and *Sncb*, with low levels of H3K27me3 in both the cortex and striatum, were expressed in the former but not the latter (Extended Data Fig. 9h). The opposite pattern was observed for genes such as *Ablim2* and *Gng7* (Extended Data Fig. 9h). *Car2*, a marker gene expressed in oligodendrocytes and abundant in the corpus callosum, presented H3K27me3 occupancy in the cortical layers, in contrast to the corpus callosum and, surprisingly, the striatum (Extended Data Fig. 9c). Thus, mechanisms other than Polycomb-mediated H3K27me3 deposition might be involved in the transcriptional repression of these genes in different areas of the CNS.

Fig. 4: Region-specific epigenetic regulation of gene expression.



a–**c**, Correlation of H3K27me3 CSS and RNA gene expression (**a**), H3K27ac GAS and RNA gene expression (**b**) and H3K4me3 GAS and RNA gene expression (**c**) in corpus callosum. **d**, Upset plot of H3K27me3 CSS and RNA gene expression in striatum and deeper and superficial cortical layers; –, low CSS or gene expression; +, high CSS or gene expression. **e**, Venn diagrams showing high (+) or low (–) CSS/GAS for different histone modifications in corpus callosum with common RNA marker genes.

Source Data

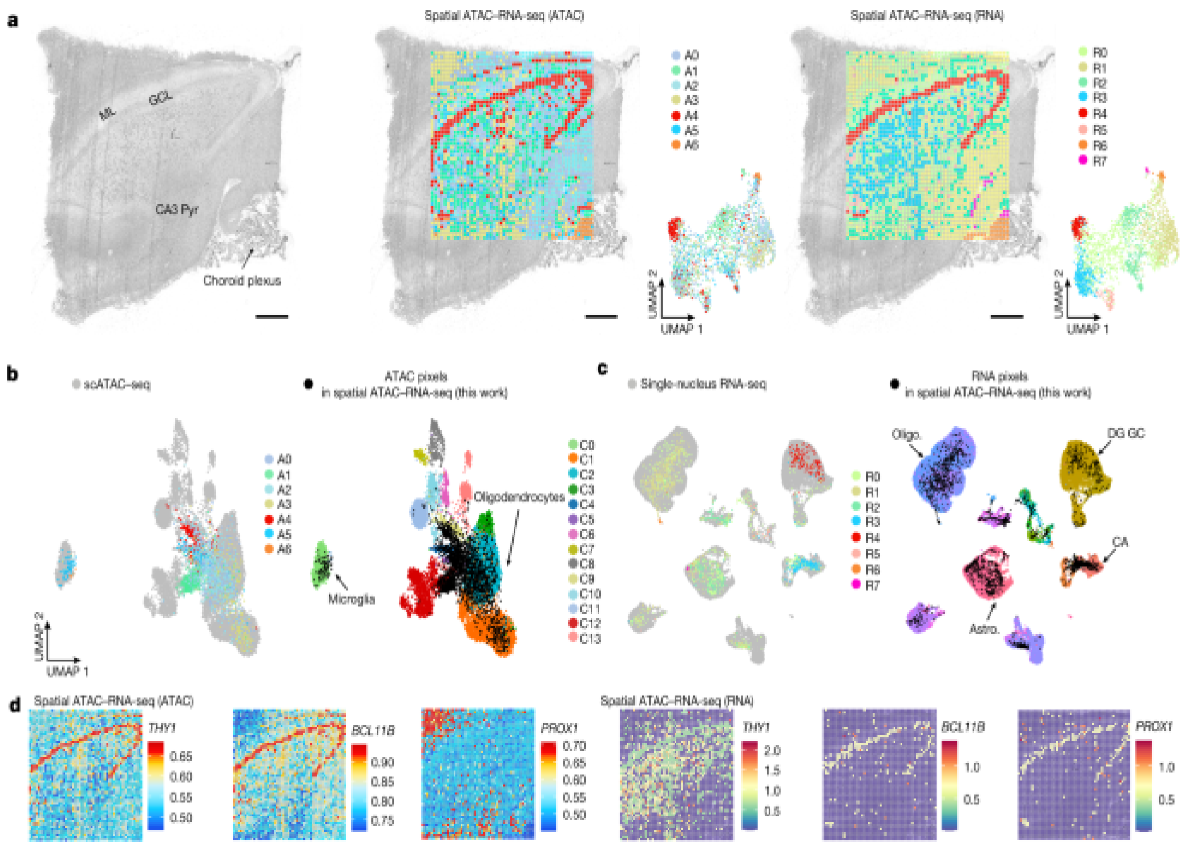
In contrast to H3K27me3, we observed a robust correlation between RNA and activating marks H3K4me3 or H3K27ac in the corpus callosum, with genes highly expressed and with high deposition in these two modalities regulating processes in oligodendroglia (Fig. 4b,c and Supplementary Fig. 12; GO for H3K27ac quadrant I and H3K4me3 quadrant I). For instance, *Mal* showed the highest activity or expression in the corpus callosum for ATAC, H3K27ac, H3K4me3 and paired RNA (Extended Data Fig. 9a).

Gpr88, a marker gene of MSN, was enriched in striatum for ATAC, H3K27ac and H3K4me3 but with low CSS in striatum for H3K27me3 (Extended Data Fig. 9f). Furthermore, we examined collective regulation among different histone modifications in the corpus callosum (Fig. 4e and Supplementary Figs. 13 and 14). In general H3K4me3 and H3K27ac showed high correlations, with their signals anticorrelated with H3K27me3 (Fig. 4e). We also observed H3K4me3 or H3K27ac and H3K27me3 co-occupancy in several gene loci in this tissue region (Fig. 4e and Supplementary Figs. 13 and 14). These genes are involved in oligodendrocyte differentiation and myelination and other processes such as protein catabolism and localization (Supplementary Figs. 13 and 14). Some of these genes, such as *Ptprz1* and *Fnbp1*, also presented higher levels of H3K27me3 and RNA (Fig. 4a). As mentioned above, this potential H3K4me3/H3K27me3 bivalency may reflect transcriptional poising.

Spatial comapping of human brain

We performed spatial ATAC–RNA-seq on adult (31-year-old male) human brain hippocampal formation, which is a complex brain region involved in cognitive functions and diseases such as major depression disease[40] and Alzheimer's disease[41]. We identified seven and eight major clusters for ATAC and RNA, respectively, and the spatial pattern agreed with major anatomical landmarks (Fig. 5a)[3], including specific marker genes (Fig. 5d). The ATAC cluster A4 represents the granule cell layer (GCL) (*THY1*, *BCL11B*) and cluster A6 corresponds to the choroid plexus (Fig. 5a,d). TF motifs (*NEUROD1* and *SNAI1*) and their spatial patterns were visualized in different tissue regions, and positive TF regulators were identified (Extended Data Fig. 10b,c). For the RNA data, we also detected distinct clusters with unique marker genes (Fig. 5a,d) such as *PROX1* and *BCL11B* enriched in cluster R4 (GCL).

Fig. 5: Spatial chromatin accessibility and transcriptome co-profiling of human hippocampus.



a, Brightfield image, spatial distribution and UMAP of all clusters based on ATAC and RNA in the human hippocampus. ML, molecular layer; Pyr, pyramidal neurons. Pixel size, 50 μm, scale bars, 1 mm. **b**, Integration of our ATAC data with scATAC-seq data[42] from human hippocampus. **c**, Integration of our RNA data with snRNA-seq data from human brain[43]. **d**, Spatial mapping of GAS and gene expression for selected marker genes in different clusters for ATAC and RNA. Oligo, oligodendrocytes; astro, astrocytes; DG, dentate gyrus.

We also conducted integration of our ATAC data and scATAC-seq data from human brain samples[42] (Fig. 5b), and our RNA data with adult human brain snRNA-seq data[43], to show the dominant cell identities and states. Cell types identified by snRNA-seq[43] were assigned to each cluster via label transfer (Fig. 5c and Extended Data Fig. 10a). Granule cells (GC) were detected in the GCL, cornu ammonis (CA) neurons were enriched in CA3–4 regions and VLMC were strongly distinguished in choroid plexus as opposed to other regions.

In general, both ATAC and RNA can readily resolve all major tissue features in this region, but spatial cosequencing of epigenome and transcriptome may provide new insights into the dynamic gene regulation mechanism that cannot be realized by single modalities. For example, *PROX1*, a signature gene defining granule neuron identity during pyramidal neuron fate selection[44], is indeed highly expressed in GCL but showed modest chromatin accessibility (Fig. 5d). This might be attributed to a minimal demand for synthesis of new *PROX1* transcripts in postmitotic mature granule cells, and is thus not required to maintain an active open chromatin state.

Discussion

Spatial omics technologies (spatial epigenomics, transcriptomics and proteomics), based on either NGS[1,2,3,4,45,46,47] or imaging[5,48], offer an unprecedented opportunity to generate new and rich insights into gene regulation in the spatial tissue context. However, to comprehensively understand the mechanism of gene regulation, different layers of omics information need to be profiled simultaneously. This was first demonstrated in dissociated single cells[6,7,8,9] but is yet to be realized directly in tissue. Imaging-based DNA sequential fluorescence in situ hybridization combined with RNA sequential fluorescence in situ hybridization detected spatial chromatin and gene expression for target genes and selected genomic loci[49]. As of today, current technologies have not been able to perform unbiased genome-wide comapping of epigenome and transcriptome on the same tissue section at the cellular level. We developed spatial ATAC–RNA-seq and spatial CUT&Tag–RNA-seq (applied to H3K27me3, H3K27ac and H3K4me3) for co-profiling of genome-wide chromatin accessibility or histone modifications in conjunction with whole transcriptome on the same tissue section at near-single-cell resolution (available for spatial browsing, see ‘Data availability’). These techniques were applied to comapping of embryonic and juvenile mouse brain, as well as adult human brain. Spatial epigenome–transcriptome cosequencing yielded two layers of spatial omics information directly in tissue, with data quality similar to that previously obtained by single modalities. These technologies allow us to examine spatiotemporal dynamics and genome-wide gene regulation mechanisms in the tissue context.

We have demonstrated either ATAC or CUT&Tag in conjunction with RNA for spatial multiomics mapping. It might be possible to combine all three—chromatin accessibility, histone modifications and transcriptome—to delineate a more comprehensive landscape of gene regulation network in tissue. It might be also possible to simultaneously measure multiple histone modifications to assess the multivalency effect on spatial gene expression regulation. Previously we demonstrated spatial profiling of transcriptome and a large panel of proteins[1]. We envision that it is possible and highly valuable to further combine epigenome, transcriptome and proteome[50] to dissect spatial patterns of cell type, state and gene regulation. Finally, spatial multiomics as reported herein may find widespread applications beyond neuroscience and developmental biology. For example, multiple modalities for spatial mapping of human disease tissues not only cross-validate one another but also better elucidate the mechanisms driving abnormal cell states that could not be readily discerned using single-modality methods. In addition, spatial information may further show how the local tissue environment affects cell state, dynamics and function across all layers of the central dogma. In this work, we developed a device with a larger mapping area and higher throughput that allows the mapping of a fourfold larger tissue area, using 100 × 100 barcodes (total 10,000 pixels, 20-µm pixel size) to cover almost the entire hemisphere of a juvenile P22 mouse brain coronal section. The serpentine channel design allows simultaneous processing of five tissue samples, each mapped for 10,000 pixels. It is feasible to further increase the mapping area by increasing the number of barcodes and further enhance throughput using automated liquid handling.

In summary, spatially resolved, genome-wide cosequencing of epigenome and transcriptome at the cellular level represents one of the most informative tools in spatial biology and can be applied to a wide range of biological and biomedical research.

Methods

Preparation of tissue slides

Mouse C57 embryo sagittal frozen sections (no. MF-104-13-C57) were purchased from Zyagen. Freshly harvested E13 mouse embryos were snap-

frozen in optimal cutting temperature compounds and sectioned at 7−10 μm thickness. Tissue sections were collected on poly-l-lysine-coated glass slides (Electron Microscopy Sciences, no. 63478-AS).

Juvenile mouse brain tissue (P21−P22) was obtained from the *Sox10:Cre-RCE:LoxP* enhanced green fluorescent protein (eGFP) line on a C57BL/6xCD1 mixed genetic background maintained at Karolinska Institutet. The line was first generated by crossing *Sox10:Cre* animals (The Jackson Laboratory, no. 025807) with *RCE:loxP* (eGFP) animals (The Jackson Laboratory, no. 032037-JAX). This was established and maintained by breeding males lacking the *Cre* allele with females carrying a hemizygous *Cre* allele; the reporter allele was maintained in homozygosity or hemizygosity in both males and females. This resulted in specific labelling of oligodendrocyte lineage with eGFP. All animals were free from mouse bacterial and viral pathogens, ectoparasites and endoparasites. The following light/dark cycle was maintained for the mice: dawn, 06:00–07:00; daylight, 07:00–18:00; dusk, 18:00–19:00; night, 19:00–06:00. Mice were housed in individually ventilated cages at a maximum number of five per cage (IVC sealsafe GM500, tecniplast). General housing parameters including temperature, ventilation and relative humidity followed the European Convention for the Protection of Vertebrate Animals used for experimental and other scientific purposes. Air quality was controlled using stand-alone air-handling units equipped with a high-efficiency particulate air filter. Relative air humidity was consistently 55 ± 10%, at a temperature of 22 °C. Husbandry parameters were monitored with ScanClime (Scanbur) units. The cages contained a card box shelter, gnawing sticks and nesting material (Scanbur), placed on hardwood bedding (TAPVEI). Mice were provided with a regular chow diet, and water was supplied by water bottle and changed weekly. Cages were changed every 2 weeks in a laminar air-flow cabinet.

Mice were sacrificed at P21/P22 (both sexes were used) by anaesthesia with ketamine (120 mg kg^{-1} body weight) and xylazine (14 mg kg^{-1} body weight), followed by transcranial perfusion with cold oxygenated artificial cerebrospinal fluid (87 mM NaCl, 2.5 mM KCl, 1.25 mM NaH_2PO_4, 26 mM $NaHCO_3$, 75 mM sucrose, 20 mM glucose, 1 mM $CaCl_2$*$2H_2O$ and 2 mM $MgSO_4$*$7H_2O$ in distilled H_2O). Following isolation, brains were

maintained for a minimal time period in artificial cerebrospinal fluid until embedding in Tissue-Tek O.C.T. compound (Sakura) and snap-freezing using a mixture of dry ice and ethanol. Coronal cryosections of 10 µm were mounted on poly-l-lysine-coated glass slides (no. 63478-AS, Electron Microscopy Sciences) or 2 × 3 square inch glass slides (AtlasXomics).

All experimental procedures were conducted following European directive no. 2010/63/EU, local Swedish directive no. L150/SJVFS/2019:9, Saknr no. L150 and Karolinska Institutet complementary guidelines for procurement and use of laboratory animals, no. Dnr. 1937/03-640. All procedures described were approved by the local committee for ethical experiments on laboratory animals in Sweden (Stockholms Norra Djurförsöksetiska nämnd, nos. 1995/2019 and 7029/2020).

Human brain tissue was obtained from the brain collection of the New York State Psychiatric Institute at Columbia University, which includes brain samples from the Republic of Macedonia. Brain tissue collection was conducted with New York State Psychiatric Institute Institutional Review Board approval, and informed consent obtained from next of kin who agreed to donate the brains and participate in psychological autopsy interviews.

We analysed brain hippocampus tissue from a 31-year-old Caucasian male individual, with no psychiatric or neurological diagnosis, who had died of a traumatic accident and had a high level of global functioning before death as measured by global assessment scale[51] score, which was 90 (scoring 1–100, with 100 the highest function), and with toxicology negative for psychotropic medications and drugs. Postmortem interval (time from demise to brain collection) was 6.5 h.

The anterior hippocampal region was dissected from a fresh-frozen coronal section (20-mm thickness) of the right brain hemisphere. The dentate gyrus region (around 10 × 10 mm^2) of the anterior hippocampal region was selected. Cryosections of 10 µm were collected on poly-l-lysine-coated glass slides (no. 63478-AS, Electron Microscopy Sciences). Samples were stored at −80 °C until further use.

Preparation of transposome

Unloaded Tn5 transposase (no. C01070010) and pA-Tn5 (no. C01070002) were purchased from Diagenode, and the transposome was assembled following the manufacturer's guidelines. The oligos applied for transposome assembly were:

Tn5ME-A, 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′,

Tn5MErev, 5′-/5Phos/CTGTCTCTTATACACATCT-3′ and

Tn5ME-B,
5′-/5Phos/CATCGGCGTACGACTAGATGTGTATAAGAGACAG-3′

DNA barcode sequences, DNA oligos and other key reagents

DNA oligos used for PCR and library construction are shown in Supplementary Table 4. All DNA barcode sequences are provided in Supplementary Tables 5 and 6 and all other chemicals and reagents in Supplementary Table 7.

Fabrication of PDMS microfluidic device

Chrome photomasks were purchased from Front Range Photomasks, with a channel width of either 20 or 50 µm. The moulds for polydimethylsiloxane (PDMS) microfluidic devices were fabricated using standard photolithography. The manufacturer's guidelines were followed to spin-coat SU-8-negative photoresist (nos. SU-2025 and SU-2010, Microchem) onto a silicon wafer (no. C04004, WaferPro). The heights of the features were about 20 and 50 µm for 20- and 50-µm-wide devices, respectively. PDMS microfluidic devices were fabricated using the SU-8 moulds. We mixed the curing and base agents in a 1:10 ratio and poured the mixture into the moulds. After degassing for 30 min the mixture was cured at 70 °C for 2 h. Solidified PDMS was extracted for further use. We have published a detailed protocol for the fabrication and preparation of the PDMS device[52].

Spatial ATAC–RNA-seq

Frozen tissue slides were thawed for 10 min at room temperature. Tissue was fixed with formaldehyde (0.2%, with 0.05 U μl^{-1} RNase Inhibitor) for 5 min and quenched with 1.25 M glycine for a further 5 min. After fixation, tissue was washed twice with 1 ml of 0.5× DPBS-RI and cleaned with deionized (DI) H_2O.

Tissue permeabilization was carried out with 200 µl of lysis buffer (3 mM $MgCl_2$, 0.01% Tween-20, 10 mM Tris-HCl pH 7.4, 0.01% NP40, 10 mM NaCl, 1% bovine serum albumin (BSA), 0.001% digitonin, 0.05 U μl^{-1} RNase inhibitor) for 15 min and washed twice with 200 µl of wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 0.1% Tween-20) for 5 min. Transposition mix (5 µl of home-made transposome, 33 µl of 1× DPBS, 50 µl of 2× Tagmentation buffer, 1 µl of 1% digitonin, 1 µl of 10% Tween-20, 0.05 U μl^{-1} RNase Inhibitor, 10 µl of nuclease-free H_2O) was added with incubation at 37 °C for 30 min. Next, 200 µl of 40 mM EDTA with 0.05 U μl^{-1} RNase inhibitor was added with incubation for 5 min at room temperature, to stop transposition. Finally, tissue sections were washed twice with 200 µl of 0.5× PBS-RI for 5 min and cleaned with DI water.

For RT, the following mixture was used: 12.5 µl of 5× RT buffer, 4.5 µl of RNase-free water, 0.4 µl of RNase inhibitor, 0.8 µl of Superase In RNase inhibitor, 3.1 µl of 10 mM deoxynucleotide triphosphate each, 6.2 µl of Maxima H Minus Reverse Transcriptase, 25 µl of 0.5× PBS-RI and 10 µl of RT primer. Tissues were incubated for 30 min at room temperature, then at 42°C for 90 min in a wet box. After the RT reaction, tissues were washed with 1× NEBuffer 3.1 and 1% RNase inhibitor for 5 min.

For first barcode (barcode A) in situ ligation, the first PDMS chip was covered to the tissue region of interest. For alignment purposes, a 10× objective (Thermo Fisher Scientific, EVOS FL Auto microscope no. AMF7000, EVOS FL Auto 2 Software Revision 2.0.2094.0) was used to take a brightfield image. The PDMS device and tissue slide were clamped tightly with a home-made acrylic clamp. Barcode A was first annealed with ligation linker 1, 10 µl of 100 µM ligation linker, 10 µl of 100 µM each barcode A and 20 µl of 2× annealing buffer (20 mM Tris pH 7.5–8.0, 100 mM NaCl, 2 mM EDTA) and mixed well. For each channel, 5 µl of ligation

master mixture was prepared with 2 μl of ligation mix (27 μl of T4 DNA ligase buffer, 72.4 μl of RNase-free water, 5.4 μl of 5% Triton X-100, 11 μl of T4 DNA ligase), 2 μl of 1× NEBuffer 3.1 and 1 μl of each annealed DNA barcode A (A1−A50/A100, 25 μM). Vacuum was used to load the ligation master mixture into 50 channels of the device, followed by incubation at 37 °C for 30 min in a wet box. The PDMS chip and clamp were removed after washing with 1× NEBuffer 3.1 for 5 min. The slide was washed with water and dried in air.

For second barcode (barcode B) in situ ligation, the second PDMS chip was covered to the slide and a further brightfield image taken with the 10× objective. An acrylic clamp was applied to clamp the PDMS and tissue slide together. Annealing of barcodes B (B1−B50/B100, 25 μM) and preparation of the ligation master mix were carried out as for barcodes A. The device was incubated at 37 °C for 30 min in a wet box. The PDMS chip and clamp were removed after washing with 1× DPBS with SUPERase In RNase inhibitor for 5 min. The slide was washed with water and dried in air. A brightfield image was then taken for further alignment.

For tissue lysis, the region of interest was digested with 100 μl of reverse crosslinking mixture (0.4 mg ml^{-1} proteinase K, 1 mM EDTA, 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1% SDS) at 58 °C for 2 h in a wet box. The lysate was collected in a 1.5 m tube and incubated at 65 °C overnight.

For DNA and cDNA separation, the lysate was purified with Zymo DNA Clean & Concentrator-5 and eluted to 100 μl of RNase-free water. The 1× B&W buffer with 0.05% Tween-20 was used to wash 40 μl of Dynabeads MyOne Streptavidin C1 beads three times. Then, 100 μl of 2× B&W buffer with 2.5 μl of SUPERase In RNase inhibitor was used to resuspend the beads, which were then mixed with the lysate and allowed to bind at room temperature for 1 h with agitation. A magnet was used to separate beads and supernatant in the lysate.

The supernatant was removed for ATAC library construction, then purified with Zymo DNA Clean & Concentrator-5 and eluted to 20 μl of RNase-free water. PCR solution (25 μl of 2× NEBNext Master Mix, 2.5 μl of 25 μM indexed i7 primer, 2.5 μl of 25 μM P5 PCR primer) was added and mixed well. PCR was first performed with the following programme: 72 °C for 5

min, 98 °C for 30 s and cycling at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min, five times. To determine additional cycles, the pre-amplified mixture (5 µl) was mixed with quantitative PCR (qPCR) solution (5 µl of 2× NEBNext Master Mix, 0.24 µl of 25× SYBR Green, 0.5 µl of 25 µM new P5 PCR primer, 3.76 µl of nuclease-free H_2O, 0.5 µl of 25 µM indexed i7 primer). The qPCR reaction was then carried out with the following programme: 98 °C for 30 s with cycling at 98 °C for 10 s, 63 °C for 30 s and 72 °C for 1 min, 20 times. The remaining pre-amplified DNA (45 µl) was amplified by running additional cycles as determined by qPCR (to reach one-third of saturated signal). The final PCR product was purified by 1× Ampure XP beads (45 µl) and eluted in 20 µl of nuclease-free H_2O.

The beads were used for cDNA library construction. They were first washed twice with 400 µl of 1× B&W buffer with 0.05% Tween-20 and once with 10 mM Tris pH 8.0 containing 0.1% Tween-20. Streptavidin beads with bound cDNA molecules were resuspended in a TSO solution (22 µl of 10 mM deoxynucleotide triphosphate each, 44 µl of 5× Maxima RT buffer, 44 µl of 20% Ficoll PM-400 solution, 88 µl of RNase-free water, 5.5 µl of 100 uM template switch primer (AAGCAGTGGTATCAACGCAGAGTGAATrGrG+G), 11 µl of Maxima H Minus Reverse Transcriptase, 5.5 µl of RNase Inhibitor). The beads were incubated at room temperature for 30 min and then at 42 °C for 90 min, with gentle shaking. After washing beads once with 400 µl of 10 mM Tris and 0.1% Tween-20 and once with water, they were resuspended in a PCR solution (110 µl of 2× Kapa HiFi HotStart Master Mix, 8.8 µl of 10 µM primers 1 and 2, 92.4 µl of RNase-free water). PCR thermocycling was carried out using the following programme: 95 °C for 3 min and cycling at 98 °C for 20 s, 65 °C for 45 s and 72 °C for 3 min, five times. After five cycles, beads were removed from the PCR solution and 25× SYBR Green was added at 1× concentration. Samples were again placed in a qPCR machine with the following thermocycling conditions: 95 °C for 3 min, cycling at 98 °C for 20 s, 65 °C for 20 s and 72 °C for 3 min, 15 times, followed by 5 min at 72 °C. The reaction was removed once the qPCR signal began to plateau. The PCR product was purified with 0.8× Ampure XP beads and eluted in 20 µl of nuclease-free H_2O.

A Nextera XT Library Prep Kit was used for library preparation. Purified cDNA (1 ng) was diluted in RNase-free water to a total volume of 5 µl, then 10 µl of Tagment DNA buffer and 5 µl of Amplicon Tagment mix were added with incubation at 55 °C for 5 min; 5 µl of NT buffer was then added, with incubation at room temperature for 5 min. PCR master solution (15 µl of PCR master mix, 1 µl of 10 µM P5 primer, 1 µl of 10 µM indexed P7 primer, 8 µl of RNase-free water) was added and the PCR reaction performed with the following programme: 95 °C for 30 s, cycling at 95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s and 72 °C for 5 min, for 12 cycles. The PCR product was purified with 0.7× Ampure XP beads to obtain the library.

An Agilent Bioanalyzer High Sensitivity Chip was used to determine size distribution and concentration of the library before sequencing. NGS was conducted on an Illumina NovaSeq 6000 sequencer (paired-end, 150-base-pair mode).

Spatial CUT&Tag–RNA-seq

The frozen tissue slide was thawed for 10 min at room temperature. Tissue was fixed with formaldehyde (0.2%, with 0.05 U μl^{-1} RNase inhibitor) for 5 min and quenched with 1.25 M glycine for a further 5 min. After fixation, the tissue was washed twice with 1 ml of wash buffer (150 mM NaCl, 20 mM HEPES pH 7.5, one tablet of protease inhibitor cocktail, 0.5 mM Spermidine) and dipped in DI water. The tissue section was permeabilized with NP40-digitonin wash buffer (0.01% digitonin, 0.01% NP40 in wash buffer) for 5 min. The primary antibody (1:50 dilution with antibody buffer (0.001% BSA, 2 mM EDTA in NP40-digitonin wash buffer) was added with incubation at 4 °C overnight. The secondary antibody (guinea pig anti-rabbit IgG, 1:50 dilution with NP40-digitonin wash buffer) was added with incubation for 30 min at room temperature. The tissue was then washed with wash buffer for 5 min. A 1:100 dilution of pA-Tn5 adaptor complex in 300-wash buffer (one tablet of Protease inhibitor cocktail, 300 mM NaCl, 0.5 mM Spermidine, 20 mM HEPES pH 7.5) was added with incubation at room temperature for 1 h, followed by a 5 min wash with 300-wash buffer. Tagmentation buffer (10 mM $MgCl_2$ in 300-wash buffer) was added with incuation at 37 °C for 1 h. Next, 40 mM EDTA with 0.05 U μl^{-1} RNase

inhibitor was added with incubation at room temperature for 5 min to stop tagmentation. Tissue was washed twice with 0.5× DPBS-RI for 5 min for further use.

For RT, two ligations and bead separation the protocols were as for spatial ATAC–RNA-seq. For construction of the CUT&Tag library, the supernatant was purified with Zymo DNA Clean & Concentrator-5 and eluted to 20 µl of RNase-free water. PCR solution (2 µl each of 10 µM P5 PCR primer and indexed i7 primer, 25 µl of NEBNext Master Mix) was added and mixed well. PCR was performed with the following programme: 58 °C for 5 min, incubation at 72 °C for 5 min and 98 °C for 30 s then cycling at 98 °C for 10 s, with incubation at 60 °C for 10 s 12 times and a final incubation at 72 °C for 1 min. The PCR product was purified by 1.3× Ampure XP beads using the standard protocol and eluted in 20 µl of nuclease-free H_2O. cDNA library construction followed the spatial ATAC–RNA-seq protocol.

An Agilent Bioanalyzer High Sensitivity Chip was used to determine size distribution and concentration of the library before sequencing. NGS was conducted on an Illumina NovaSeq 6000 sequencer (paired-end, 150-base-pair mode).

Data preprocessing

For ATAC and CUT&Tag data, linkers 1 and 2 were used to filter read 2 and sequences were converted to Cell Ranger ATAC v.1.2 format (10X Genomics). Genome sequences were in the newly formed read 1, and barcodes A and B were included in newly formed read 2. Human reference (GRCh38) or mouse reference (GRCm38) was used to align fastq files. The BED-like fragments thus obtained were used to conduct downstream analysis. The fragments file includes fragments of information on spatial locations (barcode A × barcode B) and the genome.

For RNA data, read 2 was refined to extract barcode A, barcode B and UMI. ST pipeline v.1.7.2 (ref. [53]) was used to map the processed read 1 against the mouse genome (GRCm38) or human genome (GRCh38), which created the gene matrix for downstream analysis that contains information on genes and spatial locations (barcode A × barcode B).

Data clustering and visualization

We first identified the location of pixels on tissue from the brightfield image using MATLAB 2020b (https://github.com/edicliuyang/Hiplex_proteome).

Signac v.1.8 (ref. [54]) was loaded in R v.4.1. ATAC, CUT&Tag and RNA matrices were read into Signac v.1.8 (ref. [54]). The ‘DefaultAssay’ function was used for the RNA assay. For RNA data visualization, the feature was set to 3,000 with the ‘FindVariableFeatures’ function, then data were normalized using the ‘SCTransform’ function. Normalized RNA data were clustered and RNA UMAP was built. The DefaultAssay function was applied to the ATAC/CUT&Tag assay. For ATAC/CUT&Tag data visualization, minimum cutoff was set with the ‘FindTopFeatures’ function. Data were normalized and dimensionally reduced using latent semantic indexing, then ATAC/CUT&Tag data were clustered and ATAC/CUT&Tag UMAP was built.

The DefaultAssay function was used for the joint ATAC/CUT&Tag and RNA assay. For visualization of joint ATAC/CUT&Tag and RNA data[55], the ‘FindMultiModalNeighbors’ function was used. The reduction list was set to (‘pca’, ‘lsi’), the dimensions list was set to that for RNA and ATAC/CUT&Tag, the modality weight.name was set to RNA weight and the joint UMAP was built.

To plot the above-generated UMAPs together, DefaultAssay was set to RNA and the UMAPs for ATAC/CUT&Tag, RNA or joint ATAC/CUT&Tag and RNA were visualized separately using ‘DimPlot’.

In regard to RNA spatial data visualization, the gene matrix obtained from RNA was loaded into Seurat v.4.1 (ref. [56]) as a Seurat object, and RNA metadata obtained from Signac were read into the Seurat object. All spatial maps were then plotted with the ‘SpatialPlot’ function.

In regard to ATAC/CUT&Tag spatial data visualization, the fragment file obtained from ATAC/CUT&Tag was read into ArchR v.1.0.1 (ref. [13]) as an ArchRProject and the ATAC/CUT&Tag metadata obtained from Signac were read into the ArchRProject. The data from ArchRProject were

normalized and dimensionally reduced using iterative latent semantic indexing. For GAS and CSS calculation we used the Gene Score model in ArchR. A gene score matrix was obtained for downstream analysis. The ‘getMarkerFeatures’ and ‘getMarkers’ functions in ArchR (testMethod = “Wilcoxon”, cutOff = “FDR <= 0.05”, groupBy = “seurat_cluster”) were used to find the marker genes/regions for each cluster. To visualize spatial data, results obtained from ArchR were input to Seurat v.4.1 to map the data back to the tissue. Pixel size was scaled using the ‘pt.size.factor’ parameter in the Seurat package for better visualization.

For peak-to-gene links we input RNA Seurat object using the ‘addGeneIntegrationMatrix’ function in ArchR, then peak-to-gene links were drawn with the ‘addPeak2GeneLinks’ function. Co-accessibility of peaks was calculated using the ‘addCoAccessibility’ function in ArchR.

Integrative data analysis and cell type identification

Seurat v.4.1 (ref. [56]) was used for RNA data integration and cell type identification, and the ‘SCTransform’ function to normalize our spatial RNA and scRNA-seq data. The ‘SelectIntegrationFeatures’ function was used to obtain features common to the two datasets. The ‘FindIntegrationAnchors’ function was applied to find anchors, and the ‘IntegrateData’ function to create an integrated dataset through the identified anchors. The obtained integrated dataset was clustered, showing a good match between our spatial RNA and scRNA-seq data. The ‘FindTransferAnchors’ function was used to find transfer anchors, which were then used to conduct label transfer with the ‘TransferData’ function (if more than one cell type was presented in one pixel, the major cell type was assigned).

Signac v.1.8 and Seurat v.4.1 were used for integration of our ATAC/CUT&Tag and scATAC-seq/scCUT&Tag data. The scATAC-seq/scCUT&Tag data were quantified according to our ATAC/CUT&Tag data to ensure that there were features common across both datasets. The FindIntegrationAnchors function (reduction = “rlsi”) was used to identify anchors between the two datasets. The ‘IntegrateEmbeddings’ function was used to obtain an integrated dataset through the identified anchors. The obtained integrated dataset was clustered, showing a good match between

our spatial ATAC/CUT&Tag and scATAC-seq/scCUT&Tag data. For ATAC data, the FindTransferAnchors function was used to find transfer anchors, which were then used to map scATAC-seq to our spatial ATAC data with the 'MapQuery' function.

ArchR v.1.0.1 was used for cell type identification for our ATAC/CUT&Tag data from scRNA-seq data. The gene score matrix of our ATAC/CUT&Tag was compared with the gene expression matrix from scRNA-seq, and aligned pixels from our ATAC/CUT&Tag data with cells from scRNA-seq. The function 'GeneIntegrationMatrix' was used to add pseudo-scRNA-seq profiles and cell identities.

Correlation of CSS/GAS and gene expression

Correlation analysis was performed for different tissue regions. The mouse brain hemisphere was separated into seven clusters (corpus callosum, striatum, superficial cortical layer, deeper cortical layer, lateral ventricle, lateral septal nucleus and others) according to RNA clusters and anatomical annotation, and named 'tissue_clusters'. For Fig. 4a–c, the 'FindMarkers' function (settings: min.pct = 0.25, logfc.threshold = 0.25) was used to calculate our RNA data, and genes with adjusted $P < 10^{-5}$ selected as marker genes. The getMarkerFeatures function (settings: groupBy = "tissue_clusters") was applied to calculate the GAS or CSS of genes from the marker gene list (cutOff = "FDR <= 0.05" & (cutOff = "Log2FC >= 0.1" or cutOff = "Log2FC <= −0.1")). If avg_log2FC > 0 (RNA) and Log2FC > 0 (CUT&Tag) for a specific gene, it showed in quadrant I. GO enrichment analysis was conducted with the 'enrichGO' function (qvalueCutoff = 0.05) in the clusterProfiler v.4.2 package[25].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed data reported in this paper are deposited in the Gene Expression Omnibus with accession code GSE205055. These datasets are available as web resources and can be browsed within the tissue spatial coordinates in the UCSC Cell and Genome Browser (https://brain-spatial-omics.cells.ucsc.edu), and in our own data portal generated with AtlasXplore (https://web.atlasxomics.com/visualization/Fan). Data are also available at https://ki.se/en/mbb/oligointernode. The resulting fastq files were aligned to either the human reference genome (GRCh38) (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/chromosomes/) or mouse reference genome (GRCm38) (https://hgdownload.soe.ucsc.edu/goldenPath/mm10/chromosomes/). Published data for integration and quality comparison are available online: ENCODE ATAC-seq (E13.5 mouse embryo) (https://www.encodeproject.org/search/?type=Experiment&status=released&related_series.@type=OrganismDevelopmentSeries&replicates.library.biosample.organism.scientific_name=Mus+musculus&assay_title=ATAC-seq&life_stage_age=embryonic%2013.5%20days); ENCODE RNA-seq (E13.5 mouse embryo): forebrain (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78493); midbrain (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78456); hindbrain (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78481); mouse organogenesis cell atlas (https://oncoscape.v3.sttrcancer.org/atlas.gs.washington.edu.mouse.rna/downloads); atlas of gene regulatory elements in adult mouse cerebrum (http://catlas.org/mousebrain/#!/downloads); atlas of the adolescent mouse brain (http://mousebrain.org/adolescent/downloads.html); mouse brain scCUT&Tag H3K27ac data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5949207); mouse brain scCUT&Tag H3K27me3 data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5949205); mouse brain scCUT&Tag H3K4me3 data (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163532); human hippocampus (snRNA-seq) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186538; and human hippocampus (scATAC-seq)

(https://www.ncbi.nlm.nih.gov/geo/query/a cc.cgi?acc=GSE147672). Source data are provided with this paper.

Code availability

Code for sequencing data analysis is available at Github (https://github.com/di-0579/Spatial_epigenome-transcriptome_co-sequencing) and archived at Zenodo (https://doi.org/10.5281/zenodo.7395313).

References

1. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in yissue. *Cell* **183**, 1665–1681 (2020).

2. Deng, Y. et al. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).

3. Deng, Y. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375–383 (2022).

4. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).

5. Lu, T., Ang, C. E. & Zhuang, X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **185**, 4448–4464 (2022).

6. Allaway, K. C. et al. Genetic and epigenetic coordination of cortical interneuron development. *Nature* **597**, 693–697 (2021).

7. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).

8. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).

9. Trevino, A. E. et al. Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053–5069 (2021).

10. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).

11. Ben-Chetrit, N. et al. Integration of whole transcriptome spatial profiling with protein markers. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01536-3 (2023).

12. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

13. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).

14. Gómez-López, S. et al. Sox2 and Pax6 maintain the proliferative and developmental potential of gliogenic neural stem cells In vitro. *Glia* **59**, 1588–1599 (2011).

15. Mihelec, M. et al. Novel SOX2 partner-factor domain mutation in a four-generation family. *Eur. J. Hum. Genet.* **17**, 1417–1422 (2009).

16. Chen, J. et al. A MYT1L syndrome mouse model recapitulates patient phenotypes and reveals altered brain development due to disrupted neuronal maturation. *Neuron* **109**, 3775–3792 (2021).

17. Diacou, R., Zhao, Y., Zheng, D., Cvekl, A. & Liu, W. Six3 and Six6 are jointly required for the maintenance of multipotent retinal progenitors through both positive and negative regulation. *Cell Rep.* **25**, 2510–2523 (2018).

18. Stelzer, G. et al. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinformatics* **54**,

1.30.31–31.30.33 (2016).

19. Huang, D.-F. et al. Neuronal splicing regulator RBFOX3 mediates seizures via regulating Vamp1 expression preferentially in NPY-expressing GABAergic neurons. *Proc. Natl Acad. Sci. USA* **119**, e2203632119 (2022).

20. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

21. Amador-Arjona, A. et al. SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proc. Natl Acad. Sci. USA* **112**, E1936–E1945 (2015).

22. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

23. Yoshida, H. & Goedert, M. Phosphorylation of microtubule-associated protein tau by AMPK-related kinases. *J. Neurochem.* **120**, 165–176 (2012).

24. Zhao, S. et al. Expanding the mutation and phenotype spectrum of MYH3-associated skeletal disorders. *NPJ Genom. Med.* **7**, 11 (2022).

25. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb.)* **2**, 100141 (2021).

26. Zhu, C. et al. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).

27. Meijer, M. et al. Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility. *Neuron* **110**, 1193–1210 (2022).

28. Kriegstein, A. & Alvarez-Buylla, A. The glial nature of embryonic and adult neural stem cells. *Annu. Rev. Neurosci.* **32**, 149–184 (2009).

29. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

30. Li, Y. E. et al. An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021).

31. Tasic, B. et al. Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).

32. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).

33. Kartha, V. K. et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom.* **2**, 100166 (2022).

34. Ma, C., Chitra, U., Zhang, S. & Raphael, B. J. Belayer: modeling discrete and continuous spatial variation in gene expression from spatially resolved transcriptomics. *Cell Syst.* **13**, 786–797 (2022).

35. Wu, S. J. et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* **39**, 819–824 (2021).

36. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).

37. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01535-4 (2022).

38. Marques, S. et al. Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* **352**, 1326–1329 (2016).

39. Marques, S. et al. Transcriptional convergence of oligodendrocyte lineage progenitors during development. *Dev. Cell* **46**, 504–517 (2018).

40. Otte, C. et al. Major depressive disorder. *Nat. Rev. Dis. Primers* **2**, 16065 (2016).

41. Wagner, J. et al. Medin co-aggregates with vascular amyloid-β in Alzheimer's disease. *Nature* **612**, 123–131(2022).

42. Corces, M. R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52**, 1158–1168 (2020).

43. Franjic, D. et al. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469 (2022).

44. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

45. Thornton, C. A. et al. Spatially mapped single-cell chromatin accessibility. *Nat. Commun.* **12**, 1274 (2021).

46. Cho, C.-S. et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559–3572 (2021).

47. Fu, X. et al. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* **185**, 4621–4633 (2022).

48. Fang, R. et al. Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).

49. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).

50. Chen, A. F. et al. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).

51. Endicott, J., Spitzer, R. L., Fleiss, J. L. & Cohen, J. The Global Assessment Scale: a procedure for measuring overall severity of

psychiatric disturbance. *Arch. Gen. Psychiatry* **33**, 766–771 (1976).

52. Su, G. et al. Spatial multi-omics sequencing for fixed tissue via DBiT-seq. *STAR Protoc.* **2**, 100532 (2021).

53. Navarro, J. F., Sjostrand, J., Salmen, F., Lundeberg, J. & Stahl, P. L. ST Pipeline: an automated pipeline for spatial mapping of unique transcripts. *Bioinformatics* **33**, 2591–2593 (2017).

54. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).

55. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

56. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).

Acknowledgements

We thank the Yale Center for Research Computing for guidance and use of the research computing infrastructure. The moulds for microfluidic devices were fabricated at the Yale University School of Engineering and Applied Science Nanofabrication Center. NGS was conducted at the Yale Center for Genome Analysis and at the Yale Stem Cell Center Genomics Core Facility, supported by the Connecticut Regenerative Medicine Research Fund and the Li Ka Shing Foundation. Service provided by the Genomics Core of Yale Cooperative Center of Excellence in Hematology (no. U54DK106857) was used. We thank T. Jimenez-Beristain for writing laboratory animal ethics permits and assistance with animal experiments, and the staff at Comparative Medicine-Biomedicum. M. Bartosovic was funded by the Vinnova Seal of Excellence Marie-Sklodowska Curie Actions grant RNA-centric view on oligodendrocyte lineage development (RODent). E.A. was funded by the European Union, Horizon 2020, Marie-Sklodowska Curie Actions and the grant SOLO (no. 794689). Work in G.C.-B.'s research group was supported by the Swedish Research Council (grant no. 2019-01360), the

Swedish Cancer Society (Cancerfonden, no. 190394 Pj), the Knut and Alice Wallenberg Foundation (grant nos. 2019-0107 and 2019-0089), The Swedish Society for Medical Research (grant no. JUB2019), the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, the Ming Wai Lau Centre for Reparative Medicine and Karolinska Institutet. This research was supported by Packard Fellowship for Science and Engineering (to R.F.), Yale Stem Cell Center Chen Innovation Award (to R.F.) and the US National Institutes of Health (nos. RF1MH128876, U54AG076043, U54AG079759, UG3CA257393, UH3CA257393, R01CA245313 and U54CA274509 to R.F.). Y.L. was supported by the Society for ImmunoTherapy of Cancer Fellowship.

Funding

Open access funding provided by Karolinska Institute.

Author information

Author notes

1. Yanxiang Deng

   Present address: Department of Pathology and Laboratory Medicine, Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

2. These authors contributed equally: Di Zhang, Yanxiang Deng

Authors and Affiliations

1. Department of Biomedical Engineering, Yale University, New Haven, CT, USA

   Di Zhang, Yanxiang Deng, Graham Su, Shuozhen Bao, Yang Liu & Rong Fan

2. Yale Stem Cell Center and Yale Cancer Center, Yale School of Medicine, New Haven, CT, USA

Yanxiang Deng, Graham Su, Yang Liu & Rong Fan

3. Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden

Petra Kukanja, Eneritz Agirre, Marek Bartosovic & Gonçalo Castelo-Branco

4. Department of Pathology, Yale University School of Medicine, New Haven, CT, USA

Mingze Dong, Yuval Kluger & Rong Fan

5. Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA

Mingze Dong & Yuval Kluger

6. Department of Computer Science, Princeton University, Princeton, NJ, USA

Cong Ma & Benjamin J. Raphael

7. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Sai Ma

8. Department of Biomedical Engineering, Columbia University, New York, NY, USA

Yang Xiao & Kam W. Leong

9. Department of Psychiatry, Columbia University, New York, NY, USA

Gorazd B. Rosoklija, Andrew J. Dwork, J. John Mann & Maura Boldrini

10. Division of Molecular Imaging and Neuropathology, New York State Psychiatric Institute, New York, NY, USA

Gorazd B. Rosoklija, Andrew J. Dwork, J. John Mann & Maura Boldrini

11. Macedonian Academy of Sciences & Arts, Skopje, Republic of Macedonia

Gorazd B. Rosoklija & Andrew J. Dwork

12. Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

Andrew J. Dwork

13. Department of Radiology, Columbia University, New York, NY, USA

J. John Mann

14. Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA

Kam W. Leong

15. AtlasXomics, Inc., New Haven, CT, USA

Liya Wang

16. Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA

Maximilian Haeussler

17. Applied Mathematics Program, Yale University, New Haven, CT, USA

Yuval Kluger

18. Ming Wai Lau Centre for Reparative Medicine, Stockholm Node, Karolinska Institutet, Stockholm, Sweden

Gonçalo Castelo-Branco

19. Human and Translational Immunology Program, Yale School of Medicine, New Haven, CT, USA

Rong Fan

Contributions

R.F. conceptualized the study. D.Z., Y.D. and Y.L. were responsible for the methodology. D.Z. and Y.D. carried out the experimental investigation. D.Z., Y.D., P.K., E.A., M. Bartosovic, M.D., C.M., S.M., B.J.R., Y.K., G.C.-B. and R.F. were responsible for data analysis. P.K., G.S., S.B., Y.X., K.W.L., G.B.R., A.J.D., J.J.M. and M. Boldrini were responsible for getting resources. L.W. and M.H. were responsible for the data browser. D.Z., Y.D. and R.F. wrote the original draft. All authors reviewed, edited and approved the manuscript.

Corresponding authors

Correspondence to Yanxiang Deng, Gonçalo Castelo-Branco or Rong Fan.

Ethics declarations

Competing interests

R.F., D.Z. and Y.D. are inventors of a patent provisional disclosure related to this work. R.F. is scientific founder and advisor of IsoPlexis, Singleron Biotechnologies and AtlasXomics. The interests of R.F. were reviewed and managed by Yale University Provost’s Office in accordance with the university’s conflict of interest policies. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks Steven Henikoff, Erica Scott and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Workflow of spatial-ATAC-RNA-seq and spatial-CUT&Tag-RNA-seq.

a, Schematic workflow of spatial-ATAC-RNA-seq and spatial-CUT&Tag-RNA-seq. **b–c**, Chemistry workflow of ATAC (**b**) and RNA (**c**) in spatial-ATAC-RNA-seq. **d–e**, Chemistry workflow of CUT&Tag (**d**) and RNA (**e**) in spatial-CUT&Tag-RNA-seq.

Extended Data Fig. 2 Further analysis of spatial-ATAC-RNA-seq for E13 mouse embryo.

a, H&E image from an adjacent tissue section of E13 mouse embryo. **b**, Spatial mapping of GAS and gene expression for selected marker genes in spatial-ATAC-RNA-seq. **c**, Integration of scRNA-seq data[12] from E13.5 mouse embryos with ATAC and RNA data in spatial ATAC-RNA-seq. MOCA, Mouse Organogenesis Cell Atlas. **d**, Spatial mapping of cell types identified by label transfer from scRNA-seq[12] to ATAC (top) and RNA (bottom). **e**, Genome track visualization of marker genes with peak-to-gene links for distal regulatory elements and peak co-accessibility. **f**, The expression level and the percentage of pixels in all clusters (marker genes

for each cluster) for RNA data in spatial-ATAC-RNA-seq. **g**, Dot plot showing the identification of positive TF regulators. **h**, Spatial mapping of deviation scores for selected TF motifs. **i**, Annotation of marker peaks in different clusters. **j**, GREAT enrichment analysis of marker peaks in different clusters (Binomial and hypergeometric tests).

Extended Data Fig. 3 Further pseudotime analysis of radial glia and postmitotic premature neurons in spatial-ATAC-RNA-seq.

a, GO enrichment analysis for genes from Fig. 1g. **b**,**c**, Pseudotime analysis from radial glia to postmitotic premature neurons with GAS (**b**) and gene expression (**c**). **d**, Monocle2 analyses showing different states in (**b**). **e**, Heatmap of different states along the pseudotime trajectory. **f**, GO analysis of genes in red box of (**e**) (One-sided version of Fisher's exact test, p-value was adjusted for multiple comparisons by Benjamini & Hochberg method).

Extended Data Fig. 4 Further analysis of spatial-ATAC-RNA-seq for P22 mouse brain.

a, Spatial mapping of gene activity scores and gene expression for selected marker genes in spatial-ATAC-RNA-seq. **b**, Genome track visualization of marker genes with peak-to-gene links for distal regulatory elements and peak co-accessibility. **c**, Spatial mapping of cell types identified by label transfer from scATAC-seq[30] to ATAC data. IT: intratelencephalic. PT: pyramidal tract. NP: near-projecting. CT: corticothalamic. L: layer. **d**, Spatial mapping of cell types identified by label transfer from scRNA-seq[32] to RNA data.

Extended Data Fig. 5 Further analysis of P22 mouse brain in spatial-ATAC-RNA-seq.

a, Candidate TF regulators of *Sox2*, *Pax6*, and *Mobp*. Highlighted points are TFs with abs(regulation score) ≥ 1 (−log10 scale)[33], with all other TFs shown in gray (*Z* test). **b**, Spatial mapping of deviation scores for selected TF motifs. **c**, Heatmaps of peak-to-gene links in spatial-ATAC-RNA-seq. **d**, The number of significantly correlated peaks for each gene.

Extended Data Fig. 6 Spatial chromatin accessibility and transcriptome co-sequencing of P21mouse brain.

a—**c**, Spatial distribution and UMAP of all the clusters for ATAC (**a**), RNA (**b**), and joint clustering of ATAC and RNA (**c**) in spatial-ATAC-RNA-seq for the mouse brain. Pixel size, 20 μm. Scale bar, 1 mm. **d**, Nissl-stained image from an adjacent tissue section of P21 mouse brain. Scale bar, 1 mm. **e**, Integration of ATAC data in spatial-ATAC-RNA-seq with scATAC-seq data[30] from mouse brain. **f**, Integration of RNA data in spatial-ATAC-RNA-seq with scRNA-seq data[32] from mouse brain. **g**, Spatial mapping of GAS and gene expression for selected marker genes in different clusters for ATAC and RNA in spatial-ATAC-RNA-seq.

Extended Data Fig. 7 Further analysis for spatial-CUT&Tag-RNA-seq with P22 mouse brain.

a, Integration of CUT&Tag (H3K4me3) data in spatial-CUT&Tag-RNA-seq with scCUT&Tag (H3K4me3) data[36] from mouse brain. **b**, Integration of RNA data in spatial- CUT&Tag(H3K27me3)-RNA-seq, spatial-CUT&Tag(H3K27ac)-RNA-seq, and spatial-CUT&Tag(H3K4me3)-RNA-seq with scRNA-seq data[32] from mouse brain. **c-e**, Integration of RNA data in spatial-CUT&Tag(H3K27me3)-RNA-seq (**c**), RNA data in data in spatial-CUT&Tag(H3K27ac)-RNA-seq (**d**), and RNA data in spatial-CUT&Tag(H3K4me3)-RNA-seq (**e**) with scRNA-seq data[32] from mouse brain. **f**, Spatial mapping of cell types identified by label transfer from scRNA-seq[32] to RNA data in spatial-CUT&Tag(H3K27me3)-RNA-seq. **g**, Spatial mapping of cell types identified by label transfer from scRNA-seq[32] to RNA (top) and from scRNA-seq[32] to CUT&Tag (H3K27ac, bottom) data in spatial-CUT&Tag(H3K27ac)-RNA-seq. **h**, Spatial mapping of cell types identified by label transfer from scRNA-seq[32] to RNA (top) and from scRNA-seq[32] to CUT&Tag (H3K4me3, bottom) data in spatial-CUT&Tag(H3K4me3)-RNA-seq.

Extended Data Fig. 8 Spatial epigenome and transcriptome co-sequencing and integrative analysis of P21 mouse brain.

a-c, Spatial distribution and UMAP of all the clusters for CUT&Tag (H3K27ac) (**a**), RNA (**b**), and joint clustering of CUT&Tag (H3K27ac) and RNA (**c**) in spatial-CUT&Tag-RNA-seq for the mouse brain. Pixel size, 20 µm. Scale bar, 1 mm. **d**, Nissl-stained image from an adjacent tissue section of P21 mouse brain. Scale bar, 1 mm. **e**, Integration of CUT&Tag (H3K27ac) data in spatial-CUT&Tag-RNA-seq with scCUT&Tag (H3K27ac) data[37] from mouse brain. **f**, Integration of RNA data in spatial-CUT&Tag-RNA-seq with scRNA-seq data[32] from mouse brain.

Extended Data Fig. 9 Spatial mapping of CSS, GAS, and gene expression of selected genes for P22 mouse brain.

a-g, Spatial mapping of CSS, GAS, and gene expression of *Mal* (**a**), *Mag* (**b**), *Car2* (**c**), *Grin2b* (**d**), *Syt1* (**e**), *Gpr88* (**f**), and *Ptprz1* (**g**) from ATAC and RNA in spatial-ATAC-RNA-seq, and CUT&Tag (H3K27me3, H3K27ac, or H3K4me3) and RNA in spatial-CUT&Tag-RNA-seq. **h**, Spatial mapping of CSS and gene expression of *Nav3*, *Sncb*, *Ablim2*, and *Gng7* in spatial-CUT&Tag(H3K27me3)-RNA-seq.

Extended Data Fig. 10 Further analysis of human hippocampus in spatial-ATAC-RNA-seq.

a, Spatial mapping of cell types identified by label transfer from scRNA-seq[43] to RNA data in spatial-ATAC-RNA-seq for human hippocampus. **b**, Dot plot showing the identification of positive TF regulators. **c**, Spatial mapping of deviation scores for selected TF motifs in spatial-ATAC-RNA-seq.

Supplementary information

Supplementary Information

Supplementary Figs. 1–14, Tables 2–7 and statistics and reproducibility.

Reporting Summary

Peer Review File

Supplementary Table 1

A list of cell type annotations in mouse brain.

Source data

Source Data Fig. 4.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Zhang, D., Deng, Y., Kukanja, P. *et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023). https://doi.org/10.1038/s41586-023-05795-1

- Received: 06 June 2022

- Accepted: 03 February 2023
- Published: 15 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05795-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Layering epigenomic and transcriptomic space**
  - Kirsty Minton

  Nature Reviews Genetics (2023)

Gene expression and epigenetic regulation co-mapped in brain tissues

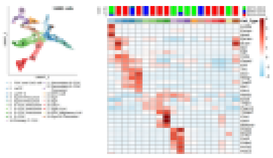
Nature Research Briefing 15 Mar 2023

Layering epigenomic and transcriptomic space

- Kirsty Minton

Nature Reviews Genetics Research Highlight 27 Mar 2023

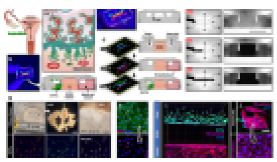
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05795-1



A differentiation roadmap of murine placentation at single-cell resolution

17 March 2023

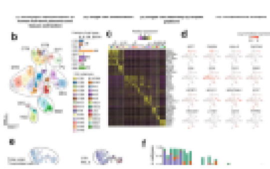
Xiangxiang Jiang, Yue Wang, … Hongmei Wang



A microphysiological model of human trophoblast invasion during implantation

15 March 2022

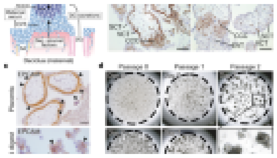
Ju Young Park, Sneha Mani, … Dan Dongeun Huh



Single-cell transcriptional profiling reveals cellular and molecular divergence in human maternal–fetal interface

28 June 2022

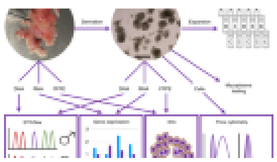
Quanlei Wang, Jinlu Li, … Zhouchun Shang



[Trophoblast organoids as a model for maternal–fetal interactions during human placentation]

28 November 2018

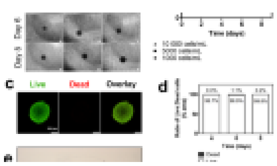
Margherita Y. Turco, Lucy Gardner, … Ashley Moffett



[Establishment and differentiation of long-term trophoblast organoid cultures from the human placenta]

09 September 2020

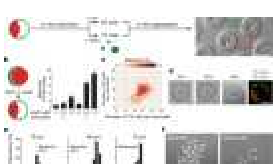
Megan A. Sheridan, Ridma C. Fernando, … Margherita Y. Turco



[Transcriptomic and functional analyses of 3D placental extravillous trophoblast spheroids]

30 August 2019

Michael K. Wong, Mishquatul Wahed, … Sandeep Raha



[Blastocyst-like structures generated solely from stem cells]

02 May 2018

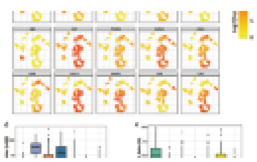
Nicolas C. Rivron, Javier Frias-Aldeguer, … Niels Geijsen



[Single-cell reconstruction of the early maternal–fetal interface in humans]

14 November 2018

Roser Vento-Tormo, Mirjana Efremova, … Sarah A. Teichmann



[Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta]

24 July 2018

Yawei Liu, Xiaoying Fan, … Hongmei Wang

- Article
- Open Access
- Published: 29 March 2023

Spatial multiomics map of trophoblast development in early

pregnancy

- Anna Arutyunyan,
- Kenny Roberts,
- Kevin Troulé,
- Frederick C. K. Wong,
- Megan A. Sheridan,
- Ilia Kats,
- Luz Garcia-Alonso,
- Britta Velten,
- Regina Hoo,
- Elias R. Ruiz-Morales,
- Carmen Sancho-Serra,
- Jarrod Shilts,
- Louis-Francois Handfield,
- Luca Marconato,
- Elizabeth Tuck,
- Lucy Gardner,
- Cecilia Icoresi Mazzeo,
- Qian Li,
- Iva Kelava,
- Gavin J. Wright,
- Elena Prigmore,
- Sarah A. Teichmann,
- Omer Ali Bayraktar ,
- Ashley Moffett ,
- Oliver Stegle ,
- Margherita Y. Turco &
- …
- Roser Vento-Tormo

Show authors

Nature **volume 616**, pages 143–151 (2023)

- 11k Accesses

- 187 Altmetric
- Metrics details

Abstract

The relationship between the human placenta—the extraembryonic organ made by the fetus, and the decidua—the mucosal layer of the uterus, is essential to nurture and protect the fetus during pregnancy. Extravillous trophoblast cells (EVTs) derived from placental villi infiltrate the decidua, transforming the maternal arteries into high-conductance vessels[1]. Defects in trophoblast invasion and arterial transformation established during early pregnancy underlie common pregnancy disorders such as pre-eclampsia[2]. Here we have generated a spatially resolved multiomics single-cell atlas of the entire human maternal–fetal interface including the myometrium, which enables us to resolve the full trajectory of trophoblast differentiation. We have used this cellular map to infer the possible transcription factors mediating EVT invasion and show that they are preserved in in vitro models of EVT differentiation from primary trophoblast organoids[3,4] and trophoblast stem cells[5]. We define the transcriptomes of the final cell states of trophoblast invasion: placental bed giant cells (fused multinucleated EVTs) and endovascular EVTs (which form plugs inside the maternal arteries). We predict the cell–cell communication events contributing to trophoblast invasion and placental bed giant cell formation, and model the dual role of interstitial EVTs and endovascular EVTs in mediating arterial transformation during early pregnancy. Together, our data provide a comprehensive analysis of postimplantation trophoblast differentiation that can be used to inform the design of experimental models of the human placenta in early pregnancy.

Main

During the nine months of human pregnancy, the successful development of the fetus is entirely dependent on its placenta. This transient extraembryonic organ is located at the interface between the mother and her fetus. Placental trophoblast cells arise from the trophectoderm surrounding the

preimplantation embryo[6]. After implantation, EVTs emerge from the cytotrophoblast shell, infiltrate the decidua—the mucosal lining of the pregnant uterus, and migrate towards the spiral arteries where they destroy the smooth muscle media. Subsequently, endovascular trophoblast cells (eEVTs) form a plug close to the cytotrophoblast shell where the arteries terminate and then eEVTs replace the endothelium[1]. In this way EVTs transform maternal arteries in the decidua basalis into high-conductance vessels[2,7,8,9]. EVTs fuse into placental bed giant cells (GCs) around the decidual–myometrial boundary and normally invade only as far as the inner third of the myometrium[10]. Placentation and successful pregnancy depend on the correct degree of trophoblast invasion, and the decidua has an important role in this process[11,12].

Our previous single-cell transcriptomics analysis of the first trimester maternal–fetal interface provided an unprecedented view of the cell states comprising this environment[13]. However, trophoblast cells present in the deeper layers of the decidua and myometrium are only present in samples of pregnant hysterectomies, and the villous syncytiotrophoblast (SCT), a multinucleated layer, is lost in classical single-cell RNA sequencing (scRNA-seq). A further difficulty is the loss of spatial context in these samples, which is essential to systematically resolve the interactions between trophoblast and decidual cells in early pregnancy. In addition, novel in vitro models have been developed recently, including trophoblast stem cells (TSCs) expanded in vitro[5] and self-renewing primary trophoblast organoids[3,4,14] (PTOs). These models can recapitulate some aspects of placental development and invasion, opening paths towards mechanistically dissecting trophoblast invasion in humans. Single-cell studies[15,16] show that these models are promising but a comprehensive benchmarking has been lacking.

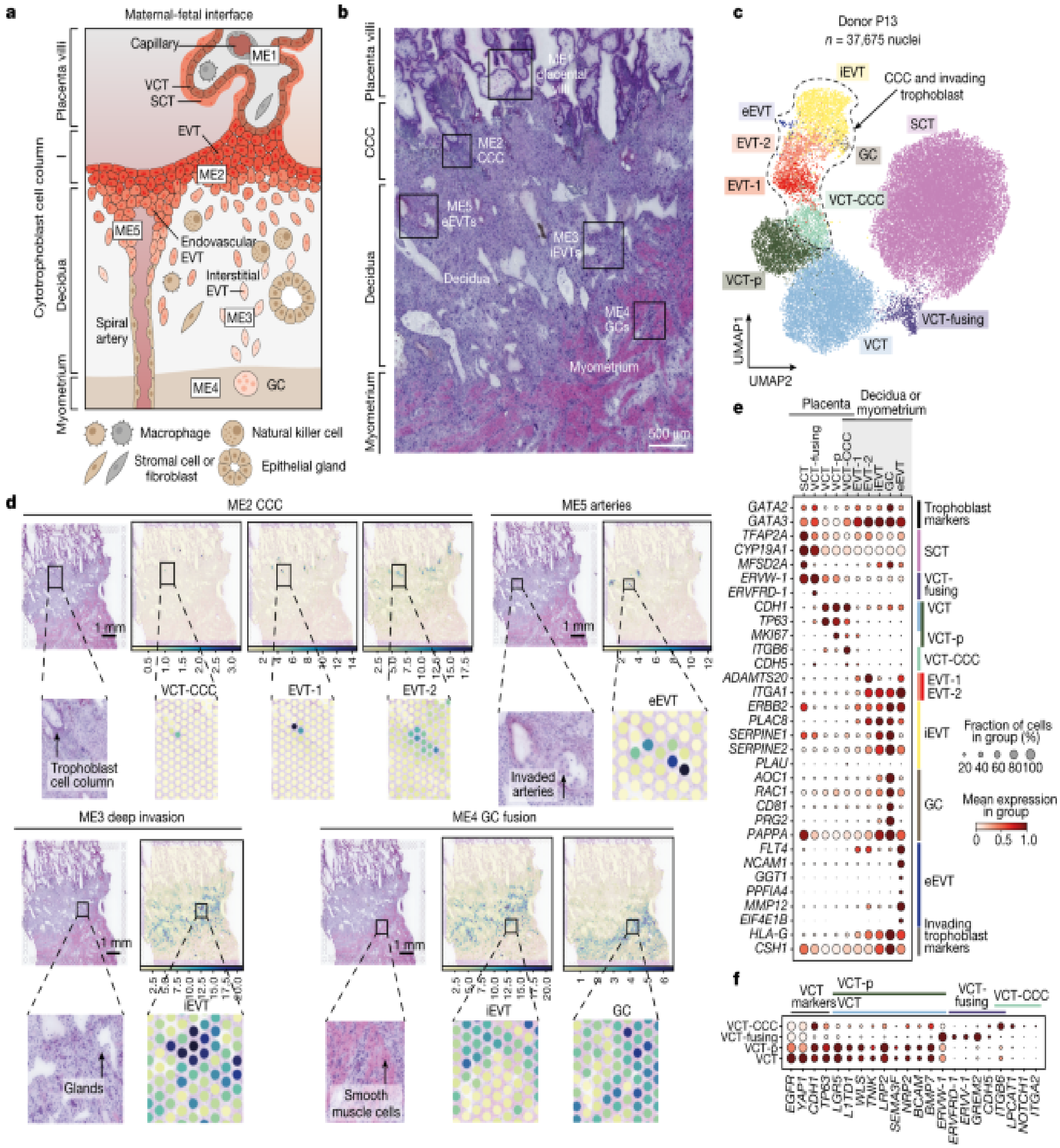
Here we present a spatially resolved single-cell multiomic characterization of the maternal–fetal interface. We examine the site of placentation from historical samples of first trimester hysterectomies, which include the entire uterus containing the placenta, decidua and myometrium. Spatiotemporal ordering of trophoblast invasion enables us to predict the potential participants regulating placentation. We use this comprehensive detailed account of trophoblast differentiation to benchmark existing PTO and TSC

models. Finally, we describe the interactions between trophoblast subsets and decidual cells that are likely to affect how arterial transformation by trophoblast occurs. Thus, we provide a description of the whole trajectory of human trophoblast cell states in the first trimester and their spatial niches.

A spatial map of human placental bed

We profiled three human implantation sites (between 6 and 9 post-conceptional weeks (PCW)) using a multimodal approach (Fig. 1a,b, Extended Data Fig. 1a,b and Supplementary Tables 1–3). Consecutive sections from frozen tissue blocks of the implantation site were used for: (1) single-nuclei RNA sequencing (snRNA-seq); (2) combined snRNA-seq and single-nuclei assay for transposase-accessible chromatin with sequencing (snATAC–seq) (we refer to the combined analysis as the multiome); and (3) spatial transcriptomics using Visium (Extended Data Figs. 1c and 2a–d). To account for the large tissue area of one donor (P13), we targeted four consecutive sections with four spatial transcriptomics capture areas (Extended Data Fig. 1d). We also profiled five decidual and three placental samples from 5–13 PCW by scRNA-seq and snRNA-seq and integrated all the data with our previous scRNA-seq dataset of the maternal–fetal interface[13] (Extended Data Figs. 1c and 2e). Our single-cell and spatial transcriptomics map is available at https://reproductivecellatlas.org.

Fig. 1: Trophoblast cell states in the early maternal–fetal interface.



a, Schematic representation of the maternal–fetal interface during the first trimester of human pregnancy. **b**, Histological overview (haematoxylin and eosin (H&E) staining) of the implantation site of donor P13 (approximately 8–9 PCW) ($n = 1$). Black outlines indicate trophoblast microenvironments in space. **c**, Uniform manifold approximation and projection (UMAP) plot of snRNA-seq of donor P13 trophoblast nuclei in the maternal–fetal interface ($n = 37{,}675$ nuclei) coloured by cell state. **d**, Overview of spatial locations of invading trophoblast cell states in Visium spatial transcriptomics data of a section of donor P13 tissue (the position of the capture area is indicated with

an arrow in Extended Data Fig. 1d). Spot colour indicates cell state density computed by cell2location, which is the number of cells of a given cell state in a Visium spot. Invading trophoblast cell states are grouped on the basis of the spatial microenvironment that they represent. **e**, Dot plots showing normalized, log-transformed and variance-scaled expression of genes (*y*-axis) characteristic of trophoblast cell states (*x*-axis) in donor P13 snRNA-seq data. **f**, Dot plots showing normalized, log-transformed and variance-scaled expression of genes (*x*-axis) characteristic of villous cytotrophoblast (*y*-axis) in donor P13 snRNA-seq data.

We examined trophoblast heterogeneity in two steps. First, we analysed the full-thickness implantation site from P13 (at around 9 PCW), as it contains both fetal (placenta) and maternal (decidua and myometrium) tissues on the same slide, and the tissue block is perfectly preserved and oriented (Fig. 1c and Extended Data Fig. 3a). Second, we validated the trophoblast populations and their markers in the integrated dataset (around 5–13 PCW) (Extended Data Fig. 3b,c). Trophoblast subsets were annotated by considering canonical markers and their spatial location (Fig. 1d–f and Extended Data Figs. 1e and 3d,e). To assign spatial coordinates we used cell2location[17], our probabilistic method to deconvolve the spatial voxels using our pre-defined snRNA-seq data. We then placed the trophoblast cells into five pre-defined microenvironments (ME1–ME5) in the tissue based on manual histological annotation.

In the placental villi (ME1), villous cytotrophoblast (VCT) fuse to form the overlying SCT layer that is in contact with maternal blood in the intervillous space. VCT subsets express high levels of *TP63* and *CDH1* in the P13 donor (Fig. 1e) and all other donors (Extended Data Fig. 3d). VCT and VCT-proliferative (VCT-p) upregulate known stem and progenitor cell markers (*LGR5*, *L1TD1* and *TP63*), Wnt signalling molecules (*WLS* and *TNIK*), the SEMA3F–NRP2 signalling complex and the VCT marker *BCAM*[18] (Fig. 1f, Extended Data Fig. 3e). We define an additional population of VCT in the placental villi that we name VCT-fusing, which the connectivity network PAGA[19] indicates is an intermediate cell state between VCT and SCT (Extended Data Fig. 3f). As VCT commit into VCT-fusing, they downregulate Wnt (*WLS*, *TNIK* and *LGR5*) and BMP signals (*BMP7* and upregulation of BMP antagonist *GREM2*), and upregulate the endogenous

retroviral genes (*ERVW-1, ERVFRD-1, ERVV-1*) known to mediate trophoblast fusion[20] (Fig. 1f and Extended Data Fig. 3e). Our strategy for isolation of nuclei enables the capture of mature multinucleated SCTs (expressing *CYP19A1* and *MFSD2A*), which were not found in previous scRNA-seq studies[13,21] (Fig. 1e and Extended Data Fig. 3d).

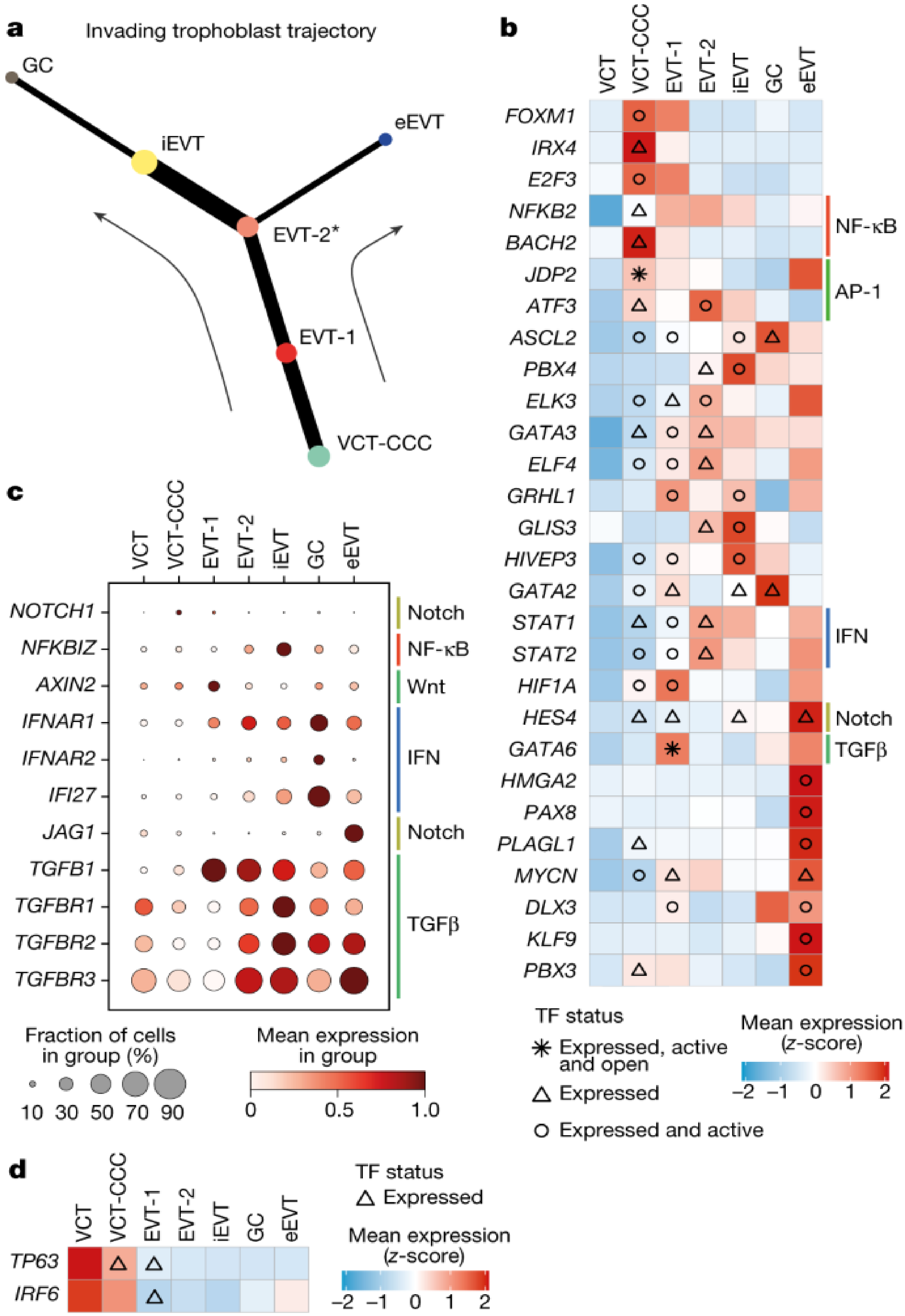
Soon after implantation, foci of cytotrophoblast cell columns (CCCs) arise from the VCTs that break through the SCT. These expand and form a shell around the conceptus that becomes discontinuous in the following weeks. EVTs begin to differentiate in cell columns but invasive EVTs emerge only when the anchoring villi attach to the maternal decidua. In the trophoblast shell (ME2), we define an additional population of CCC VCT (VCT-CCC) (Fig. 1d and Extended Data Fig. 1e). VCT-CCCs are proliferative and PAGA analysis shows they are likely to emerge from VCT or VCT-p and give rise to EVTs (Extended Data Fig. 3f). This analysis confirms that VCT is a common progenitor for both VCT-fusing, giving rise to SCT, and VCT-CCC where EVTs emerge. As they commit to become VCT-CCCs, they downregulate the Wnt pathway (*WLS*, *TNIK* and *LGR5* expression), upregulate *NOTCH1*, undergo an integrin shift (upregulating *ITGB6* and *ITGA2*), and upregulate markers characteristic of epithelial–mesenchymal transition[22] (*LPCAT1*) (Fig. 1f and Extended Data Fig. 3e). Expression of *NOTCH1* and *ITGA2* is characteristic of putative trophoblast progenitor cells located in a small niche in the CCC[23,24]. In agreement with this finding, in ME2, VCT-CCCs co-localize with EVTs (Fig. 1d and Extended Data Fig. 1e).

Trajectories of EVT defined by StOrder

To further investigate the EVT differentiation pathway as it arises from the CCCs of the anchoring villi to infiltrate maternal tissue, we leveraged both spatial and single-cell transcriptomics data using a three-step statistical framework, which we named StOrder (Extended Data Fig. 4a and Methods). First, StOrder builds a gene expression-based connectivity matrix (generated in our case by PAGA[19]) to establish putative connections between clusters (Extended Data Fig. 4b). The values in this matrix are interpreted as pairwise similarity scores for cell states in the gene expression space. Second, StOrder generates a spatial covariance matrix that reflects the

pairwise proximity of trophoblast states that co-exist in space. To do so, StOrder takes as an input the estimated cell densities per spot (derived in our case with cell2location[17]) in Visium spatial transcriptomics data, and fits a Gaussian process model that derives pairwise spatial covariance scores for all the cell state pairs (Extended Data Fig. 4a). This enables inference of which cell states are proximal in physical space and are probably gradually differentiating as they migrate. Third, StOrder reconstructs connections between cell states by combining the connectivity matrix (step 1) from single-cell transcriptomics data and the spatial covariance matrix (step 2) from the spatial data in a weighted manner (Fig. 2a and Extended Data Fig. 4a–e). In sum, StOrder reconstructs the likely cell transitions in space by taking into account both the single-cell transcriptomics and the mini-bulk spatial transcriptomics data.

Fig. 2: Transcription factors that are active during EVT invasion.



a
Invading trophoblast trajectory
GC
iEVT
eEVT
EVT-2*
EVT-1
VCT-CCC
b
VCT VCT-CCC EVT-1 EVT-2 iEVT GC eEVT
FOXM1
IRX4
E2F3
NFKB2
BACH2
JDP2
ATF3
ASCL2
PBX4
ELK3
GATA3
ELF4
GRHL1
GLIS3
HIVEP3
GATA2
STAT1
STAT2
HIF1A
HES4
GATA6
HMGA2
PAX8
PLAGL1
MYCN
DLX3
KLF9
PBX3
NF-κB
AP-1
IFN
Notch
TGFβ
TF status
Expressed, active and open
Expressed
Expressed and active
Mean expression (z-score)
−2 −1 0 1 2
c
VCT VCT-CCC EVT-1 EVT-2 iEVT GC eEVT
NOTCH1
NFKBIZ
AXIN2
IFNAR1
IFNAR2
IFI27
JAG1
TGFB1
TGFBR1
TGFBR2
TGFBR3
Notch
NF-κB
Wnt
IFN
Notch
TGFβ
Fraction of cells in group (%)
10 30 50 70 90
Mean expression in group
0 0.5 1.0
d
VCT VCT-CCC EVT-1 EVT-2 iEVT GC eEVT
TP63
IRF6
TF status
Expressed
Mean expression (z-score)
−2 −1 0 1 2

a, Representative tree of EVT differentiation trajectory inferred by StOrder (Methods). The tree shown is inferred with $\alpha = 0.4$ and $\beta = 0.5$ for snRNA-seq and spatial transcriptomics data from donors P13 (5 capture areas), P14 (2 capture areas) and Hrv43 (1 capture area). Tree edge thickness is proportional to connectivity (joint measure inferred from snRNA-seq data and spatial transcriptomics data) between two cell types connected by that edge. The asterisk indicates the bifurcation point. **b**, Heat map showing *z*-scores of normalized, log-transformed and scaled expression of transcription factor (TF) genes upregulated during trophoblast invasion in donor P13 snRNA-seq data. The *x*-axis indicates cell state, the *y*-axis lists transcription factors. Differential expression (upregulated genes) is tested along the invading trophoblast trajectory (as shown in **a**) in a retrograde manner using the limma approach (false discovery rate (FDR) < 0.05, with Bonferroni correction for multiple hypotheses testing). Coloured bars to the right of heat map indicate members of selected pathways. IFN, interferon. **c**, Dot plot showing normalized, log-transformed and variance-scaled expression of genes (*x*-axis) of signalling molecules upregulated in EVT (*y*-axis) in donor P13 snRNA-seq data. **d**, Heat map showing *z*-score of normalized, log-transformed and variance-scaled expression of transcription factors (*x*-axis) downregulated during trophoblast invasion in P13 in trophoblast states (*y*-axis). Differential expression (downregulated genes) is tested along invading trophoblast trajectory (as shown in **a**) in a retrograde manner using the limma approach (FDR < 0.05, with Bonferroni correction for multiple hypotheses testing).

StOrder enabled us to resolve the most likely trajectory for the emergence and differentiation of invasive EVTs (Fig. 2a). Consistent trajectories were obtained when reconstructing pseudotime on snRNA-seq data using Slingshot[25] (Extended Data Fig. 5a). We then calculated differentially expressed genes (DEGs) along the three trophoblast trajectories with different end points: (1) eEVT, (2) GC and (3) SCT (Extended Data Fig. 5b and Supplementary Table 5). VCT-CCCs are the precursors of EVTs-1 and EVTs-2 and co-localize with them in ME2 (Fig. 1d and Extended Data Fig. 1e). EVTs-1 are proliferative while EVTs-2 do not proliferate and upregulate the metalloprotease gene *ADAMTS20* and the integrin subunit gene *ITGA1* (Fig. 1e and Extended Data Fig. 3d). EVTs-2 are located at the distal end of the anchoring villi, and are identified as the bifurcation point between

eEVTs and interstitial EVTs (iEVTs) (Figs. 1d and 2a). Thus, EVTs-2 can transition either into iEVTs that invade through decidual stroma, or into eEVTs that move down inside the arteries.

eEVTs are present inside spiral arteries (ME5) (Fig. 1d and Extended Data Fig. 1e). Besides *NCAM1*[26,27], eEVTs also upregulate *GGT1*, *PPFIA4* and *MMP12* (Fig. 1e and Extended Data Fig. 3d). Evidence that eEVTs emerge from the distal end of the CCC is supported by their close proximity to EVTs-2 (Extended Data Fig. 6a). In our samples, we detect sporadic $NCAM1^{+}$ cells close to the cytotrophoblast shell when it is overlying a spiral artery, by single-molecule fluorescent in situ hybridisation (smFISH) (Extended Data Fig. 6b). Immunohistochemistry confirms our previous findings[26,27] that cells in the CCC do not stain with a monoclonal antibody to NCAM1, but there are scattered positive cells in the plug of eEVTs beneath this column. In a more proximal portion of the same artery all the eEVTs lining the artery are $NCAM1^{+}$ (Extended Data Fig. 6c).

Highly invasive iEVTs are found in ME3, surrounded by decidual stromal and immune cells (Fig. 1d and Extended Data Fig. 1e). iEVTs upregulate *PLAC8*[28] and plasminogen activator inhibitor genes *SERPINE1* and *SERPINE2*, with concomitant downregulation of the plasminogen activator gene *PLAU* (Fig. 1e and Extended Data Fig. 3d). iEVTs eventually fuse to form placental bed GCs deeper in the decidua and myometrium (ME4) (Fig. 1d and Extended Data Fig. 1e). GCs upregulate *RAC1* and *CD81*, and the PRG2–PAPPA complex[29] (Fig. 1d,e, Extended Data Fig. 1e and Extended Data Fig. 3c).

We next explored the regulatory programmes that might mediate EVT invasion by analysing the multimodal RNA-seq and ATAC–seq data (Extended Data Fig. 7a–c). We applied our multifactorial method MEFISTO[30] to donor P13 multimodal data, which contained the full spectra of VCT and EVT subsets (Extended Data Fig. 7d–f). MEFISTO identified 10 latent factors that jointly explain 12.5% of the variance in the RNA expression data and 3% of the chromatin accessibility data (Extended Data Fig. 7g,h). Using a logistic regression approach, we define factors 2, 4, 6 and 10 as the main driving factors of the trophoblast trajectory (Extended Data Fig. 7i–l). Factors 2, 4 and 6 explain changes along the main interstitial

trophoblast invasion pathway (VCT-CCC to GC) (Supplementary Table 4). Genes contributing strongly to these factors are *MKI67*, *CENPK* (cell cycle, factor 2); *CSF1R*, *ADAM8* and *LAIR2* (early trophoblast invasion, factor 4); *CALD1* and *COL21A1* (late trophoblast invasion, factor 6). Factor 10 captured eEVTs; the main genes contributing to this factor include *NCAM1*, *JAG1*, *ADORA1*, *EPHA1* and *HES4*.

Transcription factors in EVT subsets

To identify the major regulatory programmes driving EVT differentiation, we extracted the transcription factors that are differentially expressed and active along the EVT differentiation trajectory (Supplementary Table 6 and Methods). Activation of the *FOXM1–NOTCH1* axis is likely to lead to the differentiation of VCTs into VCT-CCCs (Fig. 2b,c and Extended Data Fig. 8a,b). Upregulation of *NOTCH1* may trigger the downregulation of *IRF6* and *TP63* expression in trophoblast[23,31] (Fig. 2d and Extended Data Fig. 8c). VCT-CCCs upregulate NF-κB pathway genes (*NFKB2* and *BACH2*) and modulate AP-1 signalling genes (*JDP2* and *ATF3*), which may result in epithelial–mesenchymal transition (Fig. 2b and Extended Data Fig. 8a). Activation of the NF-κB pathway is maintained throughout EVT differentiation (Fig. 2b and Extended Data Fig. 8a), but there is upregulation of the NF-κB inhibitor (*NFKBIZ*) at the iEVT stage (Fig. 2c and Extended Data Fig. 8b). This could be another mechanism to avoid inflammation as EVTs invade[13,32].

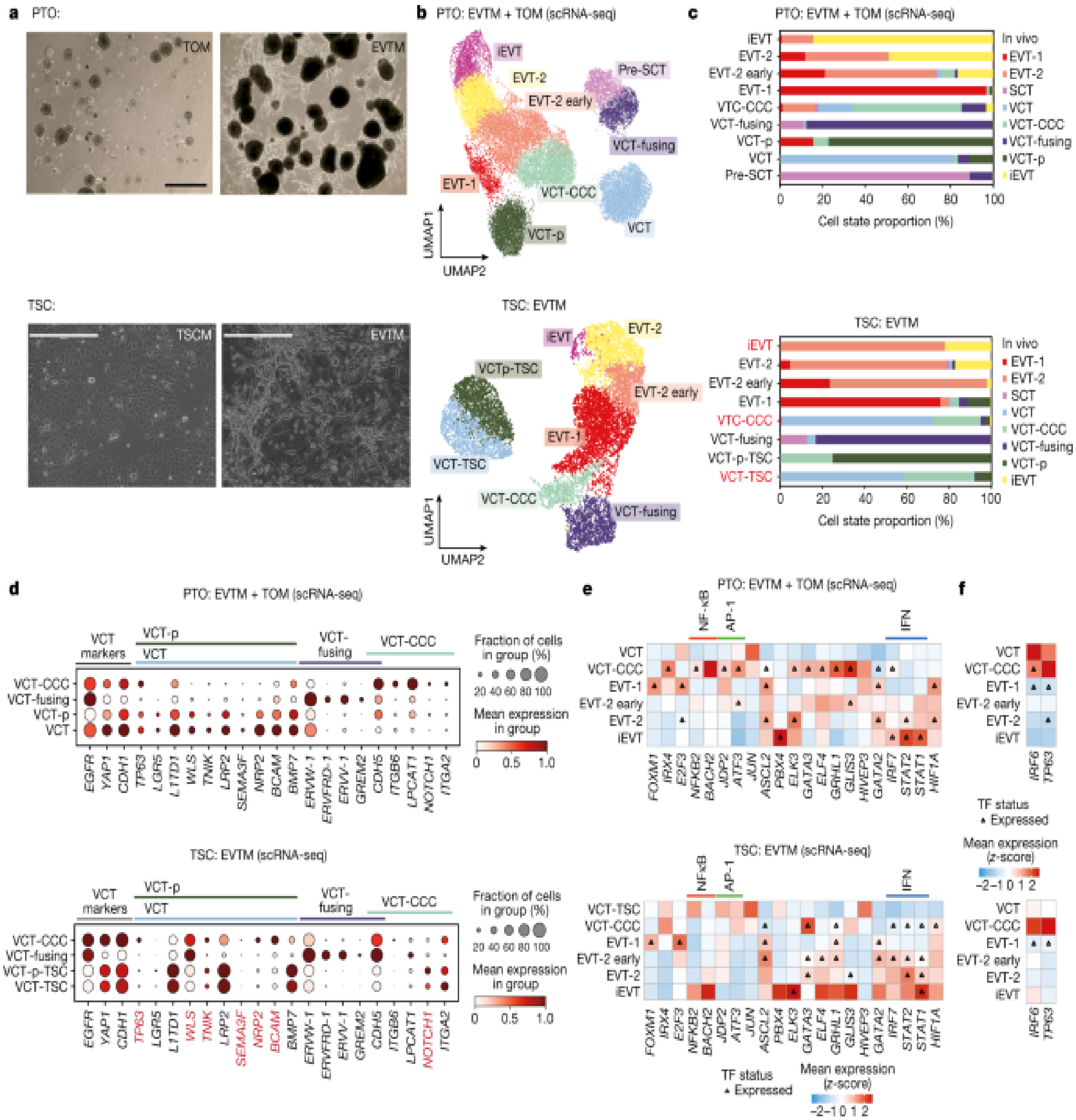
Invading EVTs intermingle with stromal and immune cells in the decidua. Decidual stromal cells secrete the Wnt inhibitor DKK1[33] and EVT invasion is characterized by inhibition of Wnt, with downregulation of the Wnt target *AXIN2* (Fig. 2c and Extended Data Fig. 8b). As they invade, iEVTs upregulate the transcription factor ASCL2[34], other transcription factors involved in cancer invasion (*ELK3–GATA3* complex[35]), as well as tumour suppressor genes (*GRHL1*) (Fig. 2b and Extended Data Fig. 8a). This is in keeping with iEVTs being non-proliferative. As iEVTs transition into GCs, they upregulate receptors of the type I interferon pathway (*IFNAR1* and *IFNAR2*) and its targets (*IFI27*) (Fig. 2c and Extended Data Fig. 8b).

The eEVTs interact with endothelial cells, which they replace, and constituents of maternal blood. eEVTs have a unique pattern of transcription factor genes, which include *HMGA2*, *PAX8*, *PLAGL1*, *MYCN* and *PBX3* (Fig. 2b and Extended Data Fig. 8a). In addition, eEVTs upregulate Notch signalling (*HES4* and *JAG1*) and the expression of TGFβ signalling genes (*TGFB1*, *TGFBR1* and *TGFBR2*) is lower than in iEVT (Fig. 2c,d and Extended Data Fig. 8a,b). *GATA6*, which is known to affect vessels by suppressing autocrine TGFβ signalling[36], is always upregulated and active in EVT-1 and maintains its high expression in eEVTs, as opposed to iEVTs. In summary, eEVT identity is marked by strong upregulation of Notch signalling and downregulation of TGFβ signalling, whereas iEVT cell fate is characterized by TGFβ upregulation and Wnt inhibition (Extended Data Fig. 8d).

Benchmarking of trophoblasts in vitro

We next explored whether the cell-intrinsic regulatory programmes that are triggered upon VCT-to-EVT differentiation are also present in EVTs derived from both self-renewing PTOs[3] and TSCs[5]. To do so, we performed scRNA-seq on: (1) PTOs differentiated in the presence of EVT medium (EVTM) (PTO-EVTM). PTOs grown in trophoblast organoid medium (TOM) (PTO-TOM) are used as controls; (2) TSCs differentiated in the presence of EVTM (TSC-EVTM). Here TSCs in trophoblast stem cell medium (TSCM) (TSC-TSCM) are used as controls (Fig. 3a and Extended Data Fig. 9a). In addition, to capture multinucleated SCT, we performed snRNA-seq on organoids grown in TOM and derived from both (3) PTOs and (4) TSCs (Extended Data Fig. 9a). We annotated the in vitro data using canonical trophoblast markers, transferring labels from the in vivo dataset into the in vitro dataset and integrating both in vivo and in vitro datasets on the same manifold (Fig. 3b and Extended Data Figs. 9b–l and 10a–e).

Fig. 3: Benchmark of EVTs derived from primary-derived trophoblast organoids and TSCs.



a, Top, phase-contrast images of PTOs plated in a Matrigel droplet and exposed to TOM or EVTM. Scale bar, 1 mm. Representative image of $n = 6$ experiments. Below, phase-contrast images of TSCs exposed to TSCM or EVTM. Scale bar, 400 µm. Representative image of $n = 2$ experiments. **b**, UMAP plot of PTO ($n = 26{,}852$ cells) and TSC (n = 9957 cells) scRNA-seq data coloured by cell state. Annotation was performed as indicated in Extended Data Fig. 9b. **c**, Bar plot representing the proportion (%) of cell states assigned to the in vitro cell states (defined by markers) using a logistic regression classifier trained on the in vivo data. Red text indicates cell states that differ between the annotations given by the logistic regression classifier

and the ones given by the expression of canonical markers. **d**, Dot plot showing normalized, log-transformed and variance-scaled expression of genes (*x*-axis) characteristic of VCT *y*-axis in PTOs (top) and TSCs (bottom). Red text indicates genes that differ from the in vivo observed expression pattern. **e**, Heat map showing *z*-scores of normalized, log-transformed and variance-scaled expression of transcription factor genes that are known to be upregulated in in vivo trophoblast invasion (see Fig. 2b). The *y*-axis indicates cell state and the *x*-axis shows transcription factor genes. **f**, Heat map showing *z*-scores of normalized, log-transformed and variance-scaled expression of transcription factor genes that are known to be downregulated in in vivo trophoblast invasion.

Projection of in vivo data onto in vitro trophoblasts using a logistic regression classifier that we trained on the in vivo dataset showed that VCT heterogeneity is better recapitulated in PTOs than in TSCs (Fig. 3c and Extended Data Fig. 10b). The four VCT subsets defined in vivo are present in PTOs and they express the same canonical markers (Fig. 3d). In PTOs, VCT-CCCs are enriched in the presence of EVTM, which triggers upregulation of the *FOXM1–NOTCH1* axis, NF-κB (*NFKB2* and *BACH2*) and AP-1 modulators (*JDP2* and *ATF3*) (Fig. 3e and Extended Data Fig. 10d). By contrast, bona fide VCTs are not found among TSCs (Fig. 3c). Instead, when grown in TSCM, cells that we call ‘trophoblast stem cells’ (VCT-TSC) are primed to become VCT-CCCs as they upregulate VCT-CCC markers (*NOTCH1* and *ITGA2*) and downregulate some of the canonical VCT markers (*TP63*, *WLS*, *TNIK*, *SEMA3F*, *NRP2* and *BCAM*) (Fig. 3d). In both TSCs and PTOs, VCT-CCCs (*CDH5*, *ITGB6* and *LCAT1*) are enriched in the presence of EVTM media, which triggers the *NOTCH–FOXM1* axis, leading to a further downregulation of *IRF6* and *TP63*[23,31] in EVT-1 (Fig. 3e,f). VCT-fusing is present in both PTOs and TSCs and accurately recapitulates its in vivo counterparts (Fig. 3c,d). snRNA-seq allowed us to capture mature SCT in PTOs (Extended Data Fig. 9d), and SCT in TSCs do not express *MFSD2A* (Extended Data Fig. 9i). Thus, our results highlight that the VCT subsets are accurately recapitulated in PTO, whereas bona fide VCTs are not found in TSCs.

VCT-CCCs in both PTOs and TSCs give rise to invasive EVT (EVT-1, EVT-2 early, EVT-2 and iEVT), whereas markers characteristic of GCs (high expression of *PRG2* and *AOC1*) and eEVT (*FLT4*, *NCAM1*, *GGT1*, *PPFIA4*,

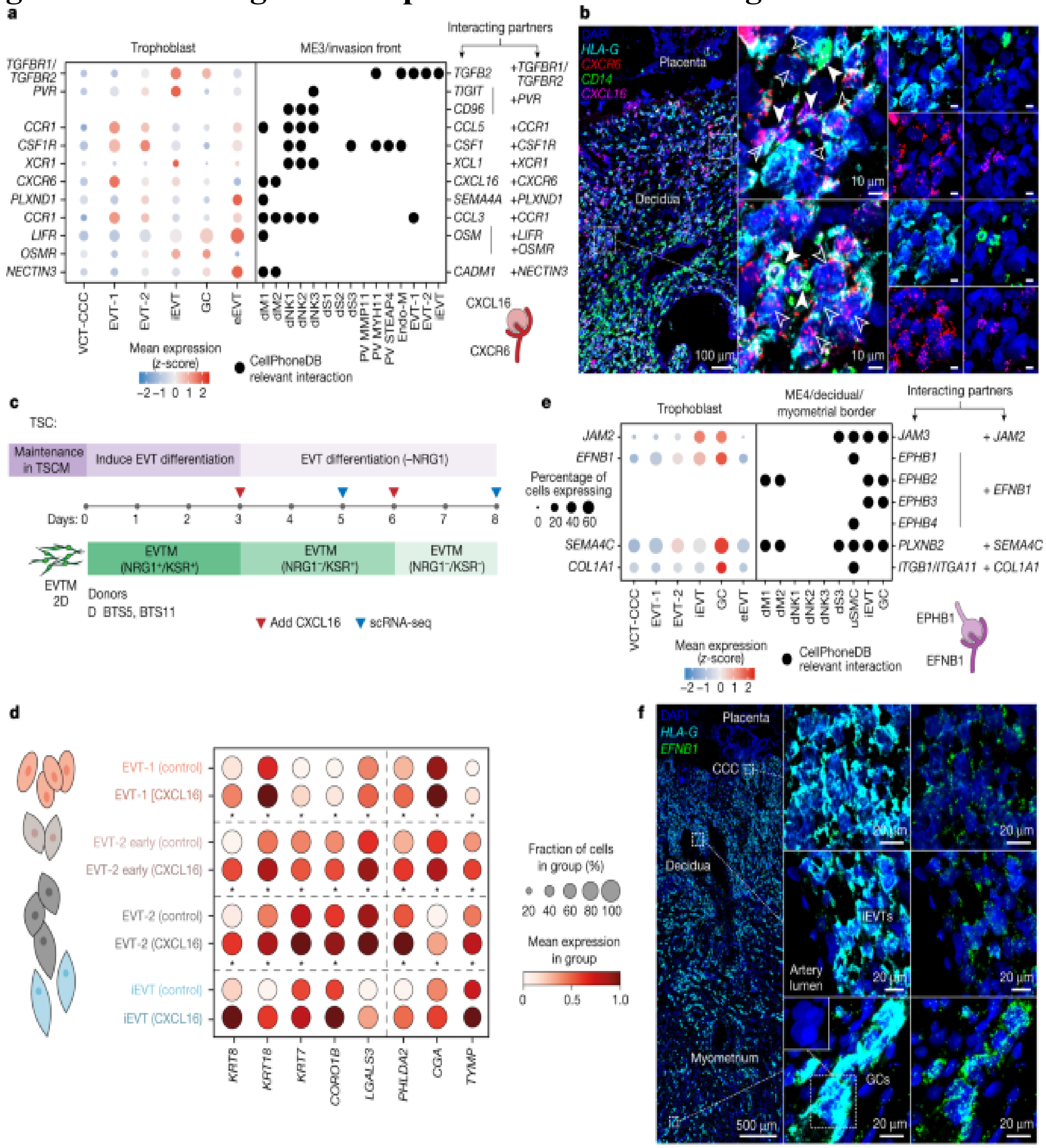
MMP12 and *EIF4E1B*) are absent in our cultures (Extended Data Fig. 10a). Despite a good representation of almost all trophoblast subsets in both in vitro models, the relative proportion and efficiency of EVT differentiation was variable (Extended Data Fig. 9g,l). Similar to in vivo EVTs, EVTs derived from PTOs or TSCs downregulate the Wnt signalling pathway (*AXIN2*), upregulate members of the TGFβ signalling pathway (*TGB1*, *TGBR1* and *TGFBR2*) and express EVT markers (*ITGA1*, *PLAC8* and *HLA-G*) (Extended Data Fig. 10a,b). Markers of deep EVT invasion (*ERBB2*, *SERPINE1*, *SERPINE2* and *PAPPA*) are upregulated in iEVTs generated in PTOs or in TSCs. However, some differences in EVT states are found between in vivo and the two in vitro trophoblast models. For PTO there is an expansion of VCT-CCCs and an early EVT-2 that upregulates markers of both VCT-CCCs (*CDH5* and *LPCAT1*) and EVTs (*CSH1*, *FBLN1*, *TIMP3*, *CD81* and *EBI3)* when compared to in vivo EVT-2 (Extended Data Fig. 10d–f). By contrast, TSC captures an early iEVT state that is assigned as EVT-2 by our logistic regression model despite upregulating iEVT markers (Fig. 3c and Extended Data Fig. 10a,c). In line with this, TSC-iEVT-early clusters together with in vivo iEVTs but expresses lower levels of invasive markers (*SERPINE2*, *PLAC8*, *HLA-G* and *RAC1*) than its in vivo counterparts (Extended Data Fig. 10d–f). Altogether, major EVT invasion programmes are conserved in both PTOs and TSCs, yet there is an expansion of an early EVT population (EVT-2 early) in PTO and a less mature iEVT-like cluster is found in TSCs. The absence of deep invasive GCs and eEVTs in these cultures suggests that maternal cues present in vivo, specifically factors from the decidual stroma or maternal arteries and blood, respectively, are essential for generating these EVT end points.

Maternal cells and EVT differentiation

We integrated single-cell and single-nuclei transcriptomics data from 18 donors to study how decidual maternal cells affect trophoblast invasion (Extended Data Figs. 1c, 2e and 11a). We used CellPhoneDB v4[37] to determine the ligand–receptor interactions that are enriched in the four decidual microenvironments (Fig. 1a and Methods). We first focused on interactions mediating trophoblast invasion (Fig. 4a). As previously described[13], decidual natural killer (dNK) cells interact with EVTs through multiple ligand–receptor pairs (PVR–TIGIT, PVR–CD96, CCR1–CCL5 and

CSF1R–CSF1). We find that the majority of these receptors are upregulated in EVT-2, near the CCCs (Fig. 4a). In this location, the CSF1–CSF1R interaction is enriched, confirming previous findings[13,38], and we reinforce this result using high-resolution multiplexed smFISH, which shows the close proximity of *CSF1*⁺ dNK cells and *CSF1R*⁺ EVT cells (Extended Data Fig. 11b).

Fig. 4: Predicted ligand–receptor interactions during EVT invasion.



a, Left, dot plot showing *z*-score of normalized, log-transformed and variance-scaled gene expression of selected receptors (*y*-axis) that are upregulated in EVT-1, EVT-2 and/or iEVT (ME3) (*x*-axis). Right, dot plot showing the presence of selected ligands (*y*-axis) in cells present in ME3 (invasion front; *x*-axis). Differential expression as in Extended Data Fig. 8a. Schematic in bottom right represents select ligand–receptor interactions. **b**, Left, high-resolution multiplexed smFISH of placenta–decidua interface showing *HLA-G* (EVT) and *CD14* (decidual macrophage), and *CXCL16* and its cognate receptor *CXCR6*. Dashed outlines indicate areas shown magnified on the right. Centre, filled and unfilled arrows indicate neighbouring *CXCL16*-expressing decidual macrophages and *CXCR6*-expressing EVTs, respectively. Images are representative of two donors. **c**, Schematic representation of the EVT differentiation experimental design, indicating time points and biological replicates in TSC models ($n = 2$ donors). **d**, Dot plot showing normalized, log-transformed and variance-scaled expression of genes (*x*-axis) that are significantly upregulated (limma, FDR < 0.05, with Bonferroni correction for multiple hypotheses testing) in the EVT subsets upon exposure to CXCL16 compared with control. **e**, Left, dot plot showing *z*-score of normalized, log-transformed and variance-scaled gene expression of selected receptors (*y*-axis) that are upregulated in GC (ME4) (*x*-axis). Right, dot plot showing the presence (*y*-axis) of selected ligands in cells present in ME4 (decidual–myometrial border; *x*-axis). Differential expression as in Extended Data Fig. 8a. Schematic in bottom right represents select ligand–receptor interactions. **f**, High-resolution multiplexed smFISH of the placenta–decidua interface showing *HLA-G* and *EFNB1*, demonstrating that expression of *EFNB1* is present throughout EVTs, including iEVTs, and higher in GCs. The inset (bottom centre) illustrates the multinucleated nature of GCs. Images representative of two donors. dM, decidual macrophages; dS, decidual stromal cells; endo-M, maternal endotheial cells; PV, perivascular.

We predicted multiple interactions between invading trophoblast cells and dM1 (*EREG*$^+$ and *IL1B*$^+$) and dM2 (*FOLR2*$^+$ and *CD14*^{high}) (Fig. 4a and Extended Data Fig. 11c,d). Maternal macrophages upregulate adhesion receptor genes, including *CADM1* (expressed in dM1 and dM2) and *SEMA4A* (expressed in dM1), whose cognate receptors *NECTIN3* and *PLXND1* are expressed in EVTs (Fig. 4a). In addition, both dM1 and dM2

express the chemokine genes *CXCL16* and *CCL3*, and their receptor genes *CXCR6*[39] and *CCR1* are upregulated in invading EVTs (Fig. 4a). *CXCR6*⁺*HLA-G*⁺ EVTs and *CXCL16*⁺*CD14*⁺ decidual macrophages are in close proximity in the implantation site (Fig. 4b). Similar to their in vivo counterparts, scRNA-seq confirms that TSC-EVTs express *CXCR6*, and we used this model to functionally validate the effect of CXCL16 on EVTs (Fig. 4c and Extended Data Fig. 11e–h). CXCL16 upregulates the expression of characteristic placental genes (*PHLDA2* and *CGA*), those involved in endothelial integrity (*TYMP*) as well as cytokeratins (*KRT7*, *KRT8* and *KRT18*), actin-binding molecules (*CORO1B*) and the galectin member *LGALS3*, previously assigned to have a role in EVT invasion[40] (Fig. 4d). This is in keeping with a role for CXCL16 in promoting trophoblast motility and function.

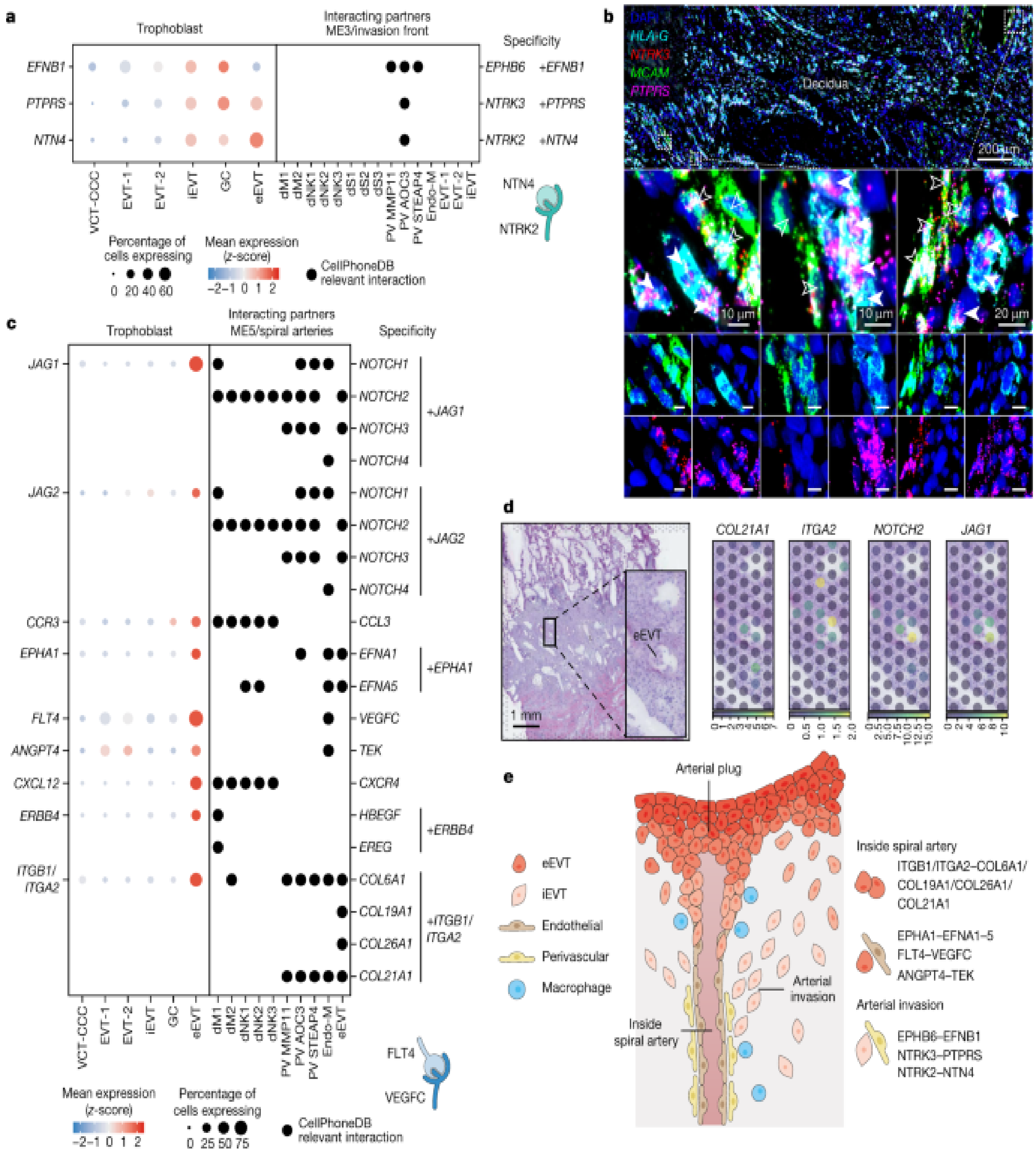
The receptors that are potentially involved in EVT invasion, including *CXCR6*, *CSF1R*[38] and *PLXND1*, are downregulated in GCs (Fig. 4a), in keeping with their presence at the limit of EVT invasion[41]. GCs form by the fusion of iEVTs and upregulate adhesion genes (*JAM2*, *EFNB1* and *SEMA4C*) whose cognate receptor genes are expressed by other iEVTs (*JAM3*, *EPHB2*, *EPHB3* and *PLXNB2*), providing potential mechanisms for fusion (Fig. 4e). A possible explanation for iEVT migration from decidua into myometrium is the specific expression of EPHB1 and EPHB4[10] by myometrial smooth muscle cells (uSMCs) which bind to EFNB1, which is upregulated in the iEVTs and GCs (Fig. 4e). We validated expression of *EFNB1* in GCs using multiplexed smFISH (Fig. 4f).

eEVT interactions with spiral arteries

Trophoblast arterial transformation during early pregnancy is crucial for pregnancy success. Initially, there is destruction of the media by iEVTs which is replaced with acellular fibrinoid material[1,27,41]. We previously defined two perivascular cell states[13], PV1 (*MCAM*-high) and PV2 (*MMP11*-high) in the arterial media. Here we combine scRNA-seq and smFISH to identify two cell states within PV1: PV1-*AOC3* (*AOC3*-high, *MYH11*-high, *FNDC1*-high and *NTRK2*-high*)* and PV1-*STEAP4 (STEAP4*-high, *EPHB6*-high and *LZTS1*-high*)* (Extended Data Fig. 12a–c). We

mapped the interactions between perivascular cell subsets and iEVT that might lead to medial destruction. Expression of *EFNB1* by iEVTs could induce their tropism towards the arteries as perivascular cells express the cognate receptor gene, *EPHB6* (Figs. 4f and 5a). We also find that iEVTs upregulate specific cell signalling molecules (*PTPRS* and *NTN4*) whose cognate receptor genes (*NTRK2* and *NTRK3*) are upregulated in PV1-*AOC3* (Fig. 5a). Neurotrophic tyrosine receptor kinases (NTRKs) can be associated with cellular survival. Whether they are involved in the 'fibrinoid change' in the arterial media[1,27,41,42] would require further exploration. Using multiplexed smFISH, we validated the close proximity between iEVTs (*HLA-G*$^+$) expressing *PTPRS* and perivascular cells (*MCAM*$^+$) expressing *NTRK2* and *NTRK3* (Fig. 5b and Extended Data Fig. 12d).

Fig. 5: Predicted ligand–receptor interactions modulating uterine arterial transformation.



a, Left, dot plot showing *z*-score of normalized, log-transformed and variance-scaled gene expression of selected receptors (*y*-axis) that are upregulated in iEVT (*x*-axis). Right, dot plot showing the presence of selected ligands (*y*-axis) in cells present in ME3 (invasion front; *x*-axis). Differential expression as in Extended Data Fig. 8a. **b**, Top, high-resolution smFISH of decidua stained for *HLA-G* and *MCAM* (PV marker), and *NTRK3* and its receptor *PTPRS*. Dashed outlines indicate areas that are shown magnified below. Middle and bottom, filled and unfilled arrows indicate

neighbouring *PTPRS*-expressing EVTs and *NTRK3*-expressing dNK cells, respectively. Images are representative of three donors. **c**, Left, dot plot showing *z*-score normalized, log-transformed and variance-scaled gene expression of selected receptors (*y*-axis) that are upregulated in eEVT (*y*-axis). In the case of a complex, the expression corresponds to the least expressed subunit of the complex (*ITGB1*). Right, dot plot showing the presence of selected ligands (*y*-axis) in cells present in ME5 (spiral arteries; *x*-axis). Differential expression as in Extended Data Fig. 8a. **d**, Overview of spatial locations of invading trophoblast cell states in Visium spatial transcriptomics data of a representative section of donor P13 tissue. The position of the capture area is indicated with an arrow in Extended Data Fig. 1d. Spot colour indicates cell state densities computed by cell2location as the number of cells of a given cell state in a Visium spot. **e**, Schematic representation of the spiral arteries in the first trimester of human pregnancy, highlighting the novel interactions between PV–iEVT, endothelial–eEVT, and eEVT–eEVT.

eEVTs initially form plugs in the spiral arteries close to the cytotrophoblast shell that limit high-pressure maternal blood flow into the intervillous space before 8–10 PCW, prior to the establishment of the haemochorial circulation[43]. eEVTs eventually replace the maternal endothelium[41,42]. Our unbiased analyses enable us to speculate how the plugs are formed. In addition to the homotypic interactions by NCAM1, eEVTs express both *ITGB1* and *ITGA2* (forming the integrin α2β1) and its cognate collagen ligands (*COL6A1*, *COL19A1*, *COL26A1* and *COL21A1*) (Fig. 5c). Active Notch signalling is suggested by upregulation of ligand (*JAG1* and *JAG2*) and receptor (*NOTCH2* and *NOTCH3*) genes (Fig. 5c). Interactions of eEVTs in the vasculature (ME5) could be mediated by *EPHA1*, *CXCL12*, *FLT4* and *ANGPT4*, with endothelial cells expressing their interacting partners *EFNA1*, *EFNA5*, *VEGFC* and *TEK* (Fig. 5c and Extended Data Fig. 12e). Using spatial transcriptomics, we visualized the expression of extracellular matrix (ECM) component (*COL21A1–ITGA2*) and Notch (*NOTCH2–JAG1*) interactions in the arterial plug (Fig. 5d).

Together, our high-resolution analyses of the spiral arteries in the decidua basalis enabled us to detect several ECM components and ligand–receptor pairs that are expressed in eEVT and maternal endothelial cells as well as in iEVT and PV subsets (Fig. 5e). These ligand–receptor interactions that occur

between maternal and fetal cells are likely to be pivotal in mediating the maternal arterial transformation that is characteristic of the first trimester of pregnancy and is essential for its success.

Discussion

In the postimplantation embryo, trophectoderm differentiates into trophoblast that invades the uterus to transform the maternal arteries. Defective trophoblast invasion is the primary underlying cause of the great obstetric syndromes that include pre-eclampsia, fetal growth restriction, unexplained stillbirth, placental abruption and preterm labour[2]. We made use of a historical collection of first-trimester pregnant hysterectomies to delineate the trophoblast landscape at the implantation site, where fetal and maternal cells intermingle. The human implantation sites profiled in our study were collected more than 30 years ago and have been stored in liquid nitrogen. We report new high-quality multiomics and spatial data, and developed a statistical framework (StOrder) that describes the complete trophoblast invasion trajectory during the first trimester of pregnancy. This includes the unbiased transcriptomics profile of eEVTs that replace the endothelium from the maternal arteries and placental bed GCs, present deeper in the decidua and the inner third of the myometrium. We use the complete in vivo trophoblast trajectory to benchmark current PTOs and TSCs in vitro trophoblast models and demonstrate that they faithfully recapitulate EVT differentiation. Terminal eEVTs and deep invasive GCs are absent in our in vitro cultures, and we reason that maternal signals from uterine cells and maternal serum are required to generate them.

Our systems biology approach has enabled us to explore potential interactions between EVTs and maternal decidual cells. First, we predict the ligand–receptor interactions between the maternal macrophages and EVT, in keeping with the importance of decidual innate immune cells for placentation[32]. We further explore the poorly described macrophage–EVT signalling axis in vitro and describe upregulation of motility genes in the EVT subsets. Second, we pinpoint the potential molecular and cellular mediators of arterial transformation during early pregnancy. Interactions between PV1-*AOC3* and iEVT could drive iEVT tropism towards the arterial wall and mediate the destruction of arterial smooth muscle media.

eEVTs have a specific ECM that could allow them to form the plug. There are also specific interactions with endothelial cells that enable eEVTs to adhere to them. These novel interactions add to our understanding of the communication between endothelial and eEVT cells[44]. The effect of defective arterial transformation in the later stages of pregnancy is well-described and underpins the great obstetric syndromes[9]. Our study increases the understanding of these major pregnancy disorders, all of which have their origins in the first trimester[45]. In addition, our roadmap of trophoblast differentiation can be used as a blueprint to design improved in vitro models that fully recapitulate the early stages of implantation.

Methods

Human samples

Placental and decidual samples used for the in vivo and in vitro profiling were obtained from elective terminations from: The MRC and Wellcome-funded Human Developmental Biology Resource (HDBR, https://www.hdbr.org), with appropriate maternal written consent and approval from the Fulham Research Ethics Committee (REC reference 18/LO/0822) and Newcastle and North Tyneside 1 Research Ethics Committee (REC reference 18/NE/0290). The HDBR is regulated by the UK Human Tissue Authority (HTA; https://www.hta.gov.uk) and operates in accordance with the relevant HTA Codes of Practice.Addenbooke's Hospital (Cambridge) under ethical approval from the Cambridge Local Research Ethics Committee (04/Q0108/23), which is incorporated into the overarching ethics permission given to the Centre for Trophoblast Research biobank for the Biology of the Human Uterus in Pregnancy and Disease Tissue Bank at the University of Cambridge under ethical approval from the East of England-Cambridge Central Research Ethics Committee (17/EE/0151) and from the London-Hampstead Research Ethics Committee (20/LO/0115).

Placental–decidual blocks (P13, P14 and P34) were collected prior to 1 September 2006 and consent for research use was not obtained. These samples are considered 'Existing Holdings' under the Human Tissue Act and as such were able to be used in this project. All the other tissue samples

used for this study were obtained with written informed consent from all participants in accordance with the guidelines in The Declaration of Helsinki 2000.

All samples profiled were histologically normal.

TSC lines BTS5 and BTS11 derived from human blastocysts by H. Okae and colleagues[5] were used in this study. Informed consent was obtained from all donors prior to the establishment of the cell line and the study was approved by the Ethics Committee of Tohoku University School of Medicine (Research license 2016-1-371), associated hospitals, the Japan Society of Obstetrics and Gynecology and the Ministry of Education, Culture, Sports, Science and Technology (Japan). This work was internally approved by HuMFre-20-0005 at the Wellcome Sanger Institute and the lines were covered by a Conditions of Use agreement with the Tohoku University School of Medicine (internal reference CG175).

Tissue cryopreservation

Fresh tissue samples of human implantation sites were embedded in cold OCT medium and flash-frozen using a dry ice-isopentane slurry as described[46].

Quality of archival frozen tissue samples was assessed by extraction of RNA from cryosections using the QIAGEN RNeasy Mini Kit, according to the manufacturer's instructions including on-column DNase I digestion. RNA quality was assayed using the Agilent RNA 6000 Nano Kit. All samples processed for Visium and single-nuclei had RIN values greater than 8.7.

Single-nuclei extraction

Single-nuclei suspensions were isolated from frozen tissue sections when performing multiomic snRNA-seq, scATAC-seq and snRNA-seq, following the manufacturer's instructions. For each OCT-embedded sample, 400 μm of tissue was prepared as 50 μm cryosections, which were paused in a tube on dry ice until subsequent processing. Nuclei were released via Dounce homogenization as described[47].

Single-cell isolation from tissue

We used the previous protocol optimized for the decidual–placental interface[13]. In short, decidual tissues were enzymatically digested in 15 ml 0.4 mg ml^{-1} collagenase V (Sigma, C9263) solution in RPMI 1640 medium (Thermo Fisher Scientific, 21875-034)/10% FCS (Biosfera, FB-1001) at 37 °C for 45 min. The supernatant was diluted with medium and passed through a 100-µm cell sieve (Corning, 431752) and then a 40-µm cell sieve (Corning, 431750). The flow-through was centrifuged and resuspended in 5 ml of red blood cell lysis buffer (Invitrogen, 00-4300) for 10 min. Placental villi were scraped from the chorionic membrane using a scalpel and the stripped membrane was discarded. The resultant villous tissue was enzymatically digested in 70 ml 0.2% trypsin 250 (Pan Biotech P10-025100P)/0.02% EDTA (Sigma E9884) in PBS with stirring at 37 °C for 9 min. The disaggregated cell suspension was diluted with medium and passed through a 100-µm cell sieve (Corning, 431752). The undigested gelatinous tissue remnant was retrieved from the gauze and further digested with 10–15 ml collagenase V at 1.0 mg ml^{-1} (Sigma C9263) in Ham's F12 medium/10% FBS with gentle shaking at 37 °C for 10 min. The disaggregated cell suspension was diluted with medium and passed through a 100 µm cell sieve (Corning, 431752). Cells obtained from both enzyme digests were pooled together and passed through a 100 µm cell sieve (Corning, 431752) and washed in Ham's F12. The flow-through was centrifuged and resuspended in 5 ml of red blood cell lysis buffer (Invitrogen, 00-4300) for 10 min.

Trophoblast in vitro cultures

Trophoblast stem cell (TSC) lines BTS5 and BTS11 derived by Okae and colleagues were grown as described previously[5]. In brief, TSC self-renewing medium (TSCM) components were substituted with local suppliers with the exception for 30% w/v BSA from WAKO Japan and CHIR99021 concentration was increased to 6 µM which maintained the undifferentiated morphology as well as preserving its EVT invasive morphology. TSCs were grown on 5 µg ml^{-1} Collagen IV (Corning) coated wells and early passaged cells between passages 24 and 26 were used for differentiation and analysis. For 2D differentiation into EVT identity, cells were seeded at a density of

1.3×10^5 per cm^2 (corresponding to 125,000 cells plated on a well of a 6-well plate) in EVTM1 detailed below supplemented with ice-cold 2% Matrigel GFR (Corning) before seeding on 1 µg ml^{-1} Collagen IV (Corning) coated wells (D0). Three days later (D3), medium was changed to EVTM2 supplemented with ice-cold 0.5% Matrigel GFR. Three days later (D6), the medium was changed to EVT medium 3 supplemented with ice-cold 0.5% Matrigel GFR. Cells were treated with TrypLE for downstream analysis 48 h later (D8). For CXCL16 induction experiments, a final concentration of 100 ng ml^{-1} CXCL16 (RnD 976-CX-025 with carrier, dissolved in 0.1%BSA(WAKO)/PBS) were supplemented to EVTM2 or EVTM3 and analysed 48 h later. The induction was controlled by supplementing an equal volume of 0.1% BSA/PBS.

In total, six trophoblast organoids were grown and differentiated into EVT as previously described[3,48]. To differentiate trophoblast organoids into EVT, organoids were cultured with TOM for ~3–4 days and transferred into EVTM1 (+NRG1) for ~4–7 days. Once trophoblasts initiate their commitment into EVT (spike emergence), EVTM2 (−NRG1) is added for 4 days. Donors were differentiated and collected in batches of three that were multiplexed on the same 10x Genomics reaction. Samples for donors 1, 2 and 3 were collected at 3 h, 24 h and 48 h after the addition of EVTM2, while samples for donors 4, 5 and 6 were collected at 48 h before, and then 0 h, 48 h and 96 h after, addition of EVTM2. Organoids grown in TOM were also collected as a control at 96h.

Media compositions have been described previously[3,5,48] and are shown here. TSCM: DMEM/F12 with Glutamax (Gibco) supplemented with 0.2% v/v FBS (Gibco), 0.3% wt/vol BSA (WAKO), 1% ITS-X (Gibco), 2.5 µg ml^{-1} l-ascorbic acid-2-phosphate (Sigma), 50 ng ml^{-1} EGF (Peprotech AF-100-15), 6 µM CHIR99021 (Tocris 4423), 0.5 µM A83-01 (Tocris 2939), 1 µM SB43154 (Tocris 1614), 0.8 mM VPA (Sigma, dissolved in H2O) and 5 µM Y-27632 (Millipore 688000). TOM: Advanced DMEM/F12, N2 supplement (at manufacturer's recommended concentration), B27 supplement minus vitamin A (at manufacturer's recommended concentration), Primocin 100 µg ml^{-1}, *N*-Acetyl-l-cysteine 1.25 mM, l-glutamine 2 mM, recombinant human EGF 50 ng ml^{-1}, CHIR99021 1.5 µM, recombinant human R-spondin-1 80 ng ml^{-1}, recombinant human FGF-2

100 ng ml^{-1}, recombinant human HGF 50 ng ml^{-1}, A83-01 500 nM, prostaglandin E2 2.5 μM, Y-27632 5 μM. EVTM1: Advanced DMEM/F12 (or DMEM/F12 for TSC-EVTM 2D), l-glutamine 2 mM, 2-mercaptoethanol 0.1 mM, penicillin/streptomycin solution 0.5% (vol/vol), BSA 0.3% (wt/vol, WAKO), ITS-X supplement 1% (vol/vol), NRG1 (Cell Signaling 5218SC) 100 ng ml^{-1}, A83-01 7.5 μM, knockout serum replacement 4% (vol/vol). EVTM2, Advanced DMEM/F12 (or DMEM/F12 for TSC-EVTM 2D), l-glutamine 2 mM, 2-mercaptoethanol 0.1 mM, penicillin/streptomycin solution 0.5% (vol/vol), BSA 0.3% (wt/vol, WAKO), ITS-X supplement 1% (vol/vol), A83-01 7.5 μM, Knockout serum replacement 4% (vol/vol) (this is the same as EVTM1 without NRG1). This medium can be stored at 4 °C for up to 1 week. EVTM3, DMEM/F12 (for TSC-EVTM 2D), l-glutamine 2 mM, 2-mercaptoethanol 0.1 mM, penicillin/streptomycin solution 0.5% (vol/vol), BSA 0.3% (wt/vol, WAKO), ITS-X supplement 1% (vol/vol), A83-01 7.5 μM (this is the same as EVTM1 without NRG1 or knockout serum replacement). This can be stored at 4 °C for up to 1 week.

H&E staining and imaging

Fresh frozen sections were removed from −80 °C storage and air dried before being fixed in 10% neutral buffered formalin for 5 min. After rinsing with deionised water, slides were stained in Mayer's haematoxylin solution for 90 s. Slides were completely rinsed in 4–5 washes of deionised water, which also served to blue the haematoxylin. Aqueous eosin (1%) was manually applied onto sections with a pipette and rinsed with deionised water after 1–3 s. Slides were dehydrated through an ethanol series (70%, 70%, 100%, 100%) and cleared twice in 100% xylene. Slides were coverslipped and allowed to air dry before being imaged on a Hamamatsu Nanozoomer 2.0HT digital slide scanner.

Multiplexed smFISH and high-resolution imaging

Large tissue section staining and fluorescent imaging were conducted largely as described previously[49]. Sections were cut from fresh frozen samples embedded in OCT at a thickness of 10–16 μm using a cryostat, placed onto SuperFrost Plus slides (VWR) and stored at −80 °C until stained. Tissue sections were processed using a Leica BOND RX to automate staining with

the RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (Advanced Cell Diagnostics, Bio-Techne), according to the manufacturers' instructions. Probes are listed in Supplementary Table 8. Prior to staining, fresh frozen sections were post-fixed in 4% paraformaldehyde in PBS for 6–8 h, then dehydrated through a series of 50%, 70%, 100%, and 100% ethanol, for 5 min each. Following manual pre-treatment, automated processing included heat-induced epitope retrieval at 95 °C for 15 min in buffer ER2 and digestion with Protease III for 15 min prior to probe hybridisation. Tyramide signal amplification with Opal 520, Opal 570, and Opal 650 (Akoya Biosciences) and TSA-biotin (TSA Plus Biotin Kit, Perkin Elmer) and streptavidin-conjugated Atto 425 (Sigma Aldrich) was used to develop RNAscope probe channels.

Stained sections were imaged with a Perkin Elmer Opera Phenix Plus High-Content Screening System, in confocal mode with 2 μm *z*-step size, using a 40× (NA 1.1, 0.149 μm/pixel) water-immersion objective. Channels: DAPI (excitation 375 nm, emission 435–480 nm), Atto 425 (excitation 425 nm, emission 463–501 nm), Opal 520 (excitation 488 nm, emission 500–550 nm), Opal 570 (excitation 561 nm, emission 570–630 nm), Opal 650 (excitation 640 nm, emission 650–760 nm).

Image stitching

Confocal image stacks were stitched as two-dimensional maximum intensity projections using proprietary Acapella scripts provided by Perkin Elmer.

10x Genomics Chromium GEX library preparation and sequencing

For the scRNA-seq experiments, cells were loaded according to the manufacturer's protocol for the Chromium Single Cell 3′ Kit v3.0, v3.1 and 5' v1.0 (10X Genomics). Library preparation was carried out according to the manufacturer's protocol to attain between 2,000 and 10,000 cells per reaction. Libraries were sequenced, aiming at a minimum coverage of 20,000 raw reads per cell, on the Illumina HiSeq 4000 or Novaseq 6000 systems using the following sequencing format: (A) read 1: 26 cycles; i7 index: 8 cycles, i5 index: 0 cycles; read 2: 98 cycles; (B) read 1: 28 cycles;

i7 index: 8 cycles, i5 index: 0 cycles; read 2: 91 cycles; (C) read 1: 28 cycles; i7 index: 10 cycles; i5 index: 10 cycles; read 2: 90 cycles (v3.1 dual).

For the multimodal snRNA-seq and scATAC-seq experiments, cells were loaded according to the manufacturer's protocol for the Chromium Single Cell Multiome ATAC + Gene Expression v1.0 to attain between 2,000 and 10,000 cells per well. Library preparation was carried out according to the manufacturer's protocol. Libraries for scATAC-seq were sequenced on Illumina NovaSeq 6000, aiming at a minimum coverage of 10,000 fragments per cell, with the following sequencing format; read 1: 50 cycles; i7 index: 8 cycles, i5 index: 16 cycles; read 2: 50 cycles.

10x Genomics Visium library preparation and sequencing

Ten-micrometre cryosections were cut and placed on Visium slides, then processed according to the manufacturer's instructions. In brief, sections were fixed with cold methanol, H&E stained and imaged on a Hamamatsu NanoZoomer S60 before permeabilization, reverse transcription and cDNA synthesis using a template-switching protocol. Second-strand cDNA was liberated from the slide and single-indexed libraries were prepared using a 10x Genomics PCR-based protocol. Libraries were sequenced (1 per lane on a HiSeq 4000), aiming for 300M raw reads per sample, with the following sequencing format; read 1: 28 cycles, i7 index: 8 cycles, i5 index: 0 cycles and read 2: 91 cycles.

Alignment and quantification of scRNA-seq and snRNA-seq data

For each sequenced single-cell and single-nucleus RNA-seq library, we performed read alignment to the 10X Genomics' GRCh38 3.0.0 human reference genome, mRNA version for scRNA-seq samples and pre-mRNA version for snRNA-seq samples, latter created following instructions from 10X Genomics: https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references#premrna. Quantification and initial quality control were performed using the Cell Ranger Software (version 3.0.2; 10X Genomics) using default parameters. Cell Ranger filtered count matrices were used for downstream analysis.

Alignment and quantification of multiome data

For each sequenced snRNA-seq and ATAC–seq (multiome) library, we performed read alignment to custom made genome consisting of 10X Genomics' GRCh38 3.0.0 pre-mRNA human reference genome and 10X Genomics Cell Ranger-Arc 1.0.1 ATAC genome, created following instructions from 10X Genomics: https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/pipelines/latest/advanced/references. Quantification and initial quality control were performed using the Cell Ranger-Arc Software (version 1.0.1; 10X Genomics) using default parameters. Cell Ranger-Arc filtered count matrices were used for downstream analysis.

Downstream scRNA-seq and snRNA-seq analysis

Detection of doublets by gene expression

We used Scrublet for cell doublet calling on a per-library basis. We used a two-step diffusion doublet identification followed by Bonferroni FDR correction and a significance threshold of 0.01, as described in[50]. Predicted doublets were not excluded from the initial analysis, but used afterwards to flag clusters with high doublet scores.

Detection of doublets by genotype

Souporcell[51] was used to deconvolute (1) maternal and fetal origin of cells and nuclei in our scRNA-seq and snRNA-seq samples (including multiome snRNA-seq); (2) assignment of cells to individuals in pooled samples (namely, samples Pla_HDBR8768477, Pla_HDBR8715512 and Pla_HDBR8715514); and (3) organoids from multiple individuals. In some samples deconvolution into maternal or fetal origin by genotype was not possible which is probably owing to the highly skewed ratio of genotypes (either extremely high (>0.95) or extremely low (<0.05) ratio of maternal to fetal droplets). In those cases, maternal–fetal origin of the cells was identified using known markers from ref. [13].

Souporcell (version 2.4.0) was installed as per instructions in https://github.com/wheaton5/souporcell and used in the following way:

```
path_to/singularity exec ./souporcell.sif souporcell_pipeline.py -i
./cellranger_path/possorted_genome_bam.bam -b
./cellranger_path/filtered_feature_bc_matrix/barcodes.tsv -f
./genome_path/genome.fa -t 8 -o souporcell_result -k 2 --skip_remap True --
common_variants ./filtered_2p_1kgenomes_GRCh38.vcf
```

Where $k = 2$ corresponds to the number of individuals to be deconvoluted (in our case either mother and fetus or pooled individuals H7 and H9 in samples Pla_HDBR8768477, Pla_HDBR8715512 and Pla_HDBR8715514. The accuracy of deconvolution was evaluated in downstream analysis once cluster identity was clear from either gene expression or predictions of logistic regression. In samples where deconvolution worked successfully, inter-individual doublets were further excluded from downstream analysis.

Filtering genes high in ambient RNA signal

To assess which genes in the scRNA-seq and snRNA-seq data were high in ambient RNA (soup) signal (further referred to as noisy genes), the following approach was undertaken separately for all the scRNA-seq and snRNA-seq samples: (1) Read in all the raw and filtered count matrices for each sample produced by Cell Ranger Software. (2) Discard droplets with < 5 unique moleular identifiers (UMIs) (likely to be fake droplets from sequencing errors). (3) Only keep data from samples which we further consider as noisy (where 'Fraction reads in cells' reported by Cell Ranger is less than 70% (guided by 10X Genomics' recommendations: https://assets.ctfassets.net/an68im79xiti/163qWiQBTVi2YLbskJphQX/e90bb82151b1cdab6d7e9b6c845e6130/CG000329_TechnicalNote_InterpretingCellRangerWebSummaryFiles_RevA.pdf). (4) Take the droplets that are in raw but are not in filtered matrices considering them as empty droplets. (5) Concatenate all raw objects with empty droplets into 1 joint raw object and do the same for filtered. (6) For all genes calculate soup probability as defined with the following equation: $P={E}_{g}^{{\rm{empty}}\,{\rm{droplets}}}/({E}_{g}^{{\rm{empty}}\,{\rm{droplets}}}+{E}_{g}^{{\rm{cells}}/{\rm{nuclei}}})$, where \

${E}_{g}^{{\rm{empty}}\;{\rm{droplets}}}$ is the total sum of expression (number of UMI counts) of gene *g* in empty droplets, and ${E}_{g}^{{\rm{cells}}/{\rm{nuclei}}}$ is the total sum of expression counts of gene *g* in droplets that are considered as cells/nuclei by Cell Ranger. (7) For all genes calculate number of cells/nuclei where the gene is detected at >0 expression level (UMI counts). (8) Label genes as noisy if their soup probability exceeds 50% quantile of soup probability distribution - done separately for cells and for nuclei.

This approach was used to estimate noisy genes in (1) donor P13 samples and (2) all donors' samples. Donor P13 noisy genes were excluded during mapping onto space (Visium, see 'Location of cell types in Visium data'), whereas all donors' noisy genes (labelled using nuclei-only derived threshold in step 8 to not over-filter genes based on the higher quality portion of the data which in our case in scRNA-seq) were excluded during all donors analysis of the whole atlas of all the cell states at the maternal–fetal interface.

Quality filters, alignment of data across different batches and clustering

We integrated the filtered count matrices from Cell Ranger and analysed them with scanpy (version 1.7.1), with the pipeline following their recommended standard practises. In brief, we excluded genes expressed by less than three cells, excluded cells expressing fewer than 200 genes, and cells with more than 20% mitochondrial content. After converting the expression space to log(CPM/100 + 1), the object was transposed to gene space to identify cell cycling genes in a data-driven manner, as described in[50,52]. After performing principal component analysis (PCA), neighbour identification and Louvain clustering, the members of the gene cluster including known cycling genes (*CDK1*, *MKI67*, *CCNB2* and *PCNA*) were flagged as the data-derived cell cycling genes, and discarded in each downstream analysis where applicable.

Next, to have an estimate of the optimal number of latent variables to be used later in the single-cell variational inference (scVI) workflow for dimensionality reduction and batch correction, we identified highly variable genes, scaled the data and calculated PCA to observe the variance ratio plot

and decide on an elbow point which defined values of n_latent parameter which were then used to correct for batch effect by 10X library batch ('sample') with scVI. Number of layers in scVI models was tuned manually to allow for better integration. The resulting latent representation of the data was used for calculating neighbourhood graph, UMAP and further Louvain clustering. For trophoblast organoid scRNA-seq and snRNA-seq, data were integrated with Harmony by donor using theta = 0 parameter.

Analysis was done separately for (a) donor P13 trophoblast compartment and (b) all donors' data (all cell states). In both analyses (a) and (b) trophoblast data was analysed separately with consecutive rounds of re-analysis upon exclusion of clusters of noisy nature (exhibiting gene expression characteristic of more than 1 distinct population). In addition, in all donors' analysis fibroblast (maternal and fetal separately) and maternal NK, T, myeloid, epithelial, endothelial and perivascular compartments were reanalysed separately using the approach described in the previous paragraph to achieve fine grain annotation.

Differential gene expression analysis

Differential gene expression analysis was performed with limma (limma version 3.46.0, edgeR version 3.32.1) with "cell_or_nucleus" covariate (scRNA-seq or snRNA-seq (including multiome snRNA-seq) origin of each droplet) backwards along the trajectory that was derived using stOrder approach, namely for the following 6 comparisons: VCT-CCC vs VCT (VCT and VCT-p cell states together); EVT-1 vs VCT-CCC; EVT-2 vs EVT-1; iEVT vs EVT-2; GC vs iEVT; eEVT vs EVT-2. Only significant DEGs were considered for downstream analysis, namely those with FDR (bonferroni) < 0.05).

Alignment, quantification and quality control of multiome ATAC data

We processed scATAC-seq libraries coming from multiome samples (read filtering, alignment, barcode counting, and cell calling) with 10X Genomics Cell Ranger-Arc (version 1.0.1) using the pre-built 10X GRCh38 genome (version corresponding to Cellranger-arc 1.0.1) as reference. We called the

peaks using an in-house implementation of the approach described in Cusanovich et al. [53] (available at https://github.com/cellgeni/cellatac, revision 21-099). In short, the genome was broken into 5-kb windows and then each cell barcode was scored for insertions in each window, generating a binary matrix of windows by cells. Matrices from all samples were concatenated into a unified matrix, which was filtered to retain only the top 200,000 most commonly used windows per sample. Using Signac (https://satijalab.org/signac/ version 0.2.5), the binary matrix was normalized with term frequency-inverse document frequency (TF-IDF) followed by a dimensionality reduction step using Singular Value Decomposition (SVD). The first latent semantic indexing (LSI) component was ignored as it usually correlates with sequencing depth (technical variation) rather than a biological variation[53]. The 2–30 top remaining components were used to perform graph-based Louvain clustering. Next, peaks were called separately on each cluster using macs2[54]. Finally, peaks from all clusters were merged into a master peak set (that is, peaks overlapping in at least one base pair were aggregated) and used to generate a binary peak-by-cell matrix, indicating any reads occurring in each peak for each cell.

This analysis was done separately for (1) all multiome data at first and (2) trophoblast-only subset of the multiome data. In the latter analysis we used annotation labels from the RNA counterpart of the multiome samples to perform peak calling.

Alignment, quantification and quality control of Visium data

For each 10X Genomics Visium sample, we used Space Ranger Software Suite (version 1.1.0) to align to the GRCh38 human reference pre-mRNA genome (official Cell Ranger reference, version 3.0.0) and quantify gene counts. Spots were automatically aligned to the paired H&E images by Space Ranger software. All spots under tissue detected by Space Ranger were included in downstream analysis.

Downstream analysis of 10X Genomics Visium data

Location of cell types in Visium data

To locate the cell states in the Visium transcriptomics slides, we used the cell2location tool v0.06-alpha[55]. As reference, we used snRNA-seq data of donor P13. We used general cell state annotations from the joint all donors' analysis (corresponding to donor P13 data), with the exception of the trophoblast lineage. Trophoblast annotations were taken from donor P13-only analysis of the trophoblast compartment. Using information about which genes are noisy (high in ambient RNA signal) in donor P13 snRNA-seq data (details in 'Filtering genes high in ambient RNA signal'), we excluded those from the reference and Visium objects prior to cell2location model training which significantly improved the results of mapping (namely, eliminated off-target mapping of cell states—that is, made results of mapping more specific to the correct anatomical regions). Following the tutorial at https://cell2location.readthedocs.io/en/latest/notebooks/cell2location_tutorial.html#Cell2location:-spatial-mapping, we trained cell2location model with default parameters using 10X library as a batch covariate in the step of estimation of reference cell-type signatures. Results were visualized with scanpy (version 1.7.1). Plots represent estimated abundance of cell types (cell densities) in Visium spots.

Subsetting Visium data into anatomical regions with SpatialDE2

We used SpatialDE2[56] tissue segmentation algorithm to assign Visium spots to three anatomical regions: (1) placenta; (2) decidua and villi tips; and (3) myometrium. We used mRNA abundances from the deconvolution results obtained with cell2location[17] in SpatialDE2 tissue segmentation. Assignment of obtained Visium spot clusters to regions was done upon visual inspection. Locations of certain fibroblast cell states indicative of the specific anatomical region (uterine smooth muscle cells, uSMC and dS cell states) were also used to guide this assignment. In addition, low-quality spots were discarded on the basis of not being under tissue and having low count and gene coverage (visual inspection).

For more details, please refer to the following notebook: https://github.com/ventolab/MFI/blob/main/2_inv_troph_trajectory_and_TFs/2-

1_stOrder_inv_troph/S1_regions_analysis_for_SpCov_model_and_later_for_CellPhone.ipynb

Downstream snATAC-seq analysis

Quality filters

To obtain a set of high-quality peaks for downstream analysis, we filtered out peaks that (1) were included in the ENCODE blacklist, (2) have a width outside the 210–1,500 bp range, and (3) were accessible in less than 5% of cells from a cellatac cluster. Low-quality cells were also removed by setting to 4 the minimum threshold for log1p-transformed total counts per cell.

Alignment of data across different batches and clustering

We adopted the cisTopic approach[57,58] for the core of our downstream analysis. cisTopic employs latent Dirichlet allocation (LDA) to estimate the probability of a region belonging to a regulatory topic (region–topic distribution) and the contribution of a topic within each cell (topic–cell distribution). The topic–cell matrix was used for constructing the neighbourhood graph, computing UMAP projections and clustering with the Louvain algorithm. After this was done for all cell states, clusters corresponding to trophoblast cell states (based on the unbiased clustering done here and annotation labels coming from the RNA counterpart of this multiome data) were further subsetted and reanalysed following the same pipeline.

Gene activity scores

Next, we generated a denoised accessibility matrix (predictive distribution) by multiplying the topic–cell and region–topic distribution and used it to calculate gene activity scores. To be able to integrate them with scRNA-seq and snRNA-seq data, gene activity scores were rounded and multiplied by a factor of 10^7, as described[58].

Cell-type annotation of invading trophoblast

Final labels of invading trophoblast in snATAC-seq data were directly transferred from RNA counterpart of the multiome data.

Join inference of trophoblast invasion from gene expression and spatial data

StOrder is a computational framework for joint inference of cellular differentiation trajectories from gene expression data and information about location of cell states in physical space (further referred to as spatial data).

It consists of three principal steps:

1. 1.

   Calculate pairwise cell state connectivity from gene expression data (here we use snRNA-seq data).

2. 2.

   Calculate pairwise cell state proximity in physical space from spatial data (here we use Visium spatial transcriptomics data) using a new spatial covariance model.

3. 3.

   Combine connectivity matrices from steps 1 and 2 in a weighted expression to reconstruct the putative tree structure of the differentiation trajectory.

First, StOrder relies on a gene expression-based connectivity matrix (generated in our case by PAGA[59]) that establishes potential connections between cell state clusters defined by single-cell or single-nucleus transcriptomics datasets. The values in this matrix can be interpreted as pairwise similarity scores for cell states in gene expression space. In our case we used snRNA-seq data from P13 as it contains all trophoblast subsets.

Second, StOrder generates a spatial covariance matrix that reflects pairwise proximity of cell states that co-exist in space and smoothly transition from one state to another while physically migrating in space. To do so, StOrder takes as an input the deconvolution results (derived in our case with cell2location[17]) of Visium spatial transcriptomics data. Here, we used all spatial transcriptomics data profiled (donors P13, P14 and Hrv43). Then, it fits a Gaussian process model that derives pairwise spatial covariance scores for all the cell state pairs with the following model:

$${\rm{vec}}({{\bf{Y}}}_{i},{{\bf{Y}}}_{j}) \sim {\mathscr{N}}\,\left(0,\,\left(\begin{array}{cc}{\sigma }_{1}^{(1)} & {\sigma }_{2}^{(1)}\\ {\sigma }_{2}^{(1)} & {\sigma }_{3}^{(1)}\end{array}\right)\otimes K({\bf{X}},l)+\left(\begin{array}{cc}{\sigma }_{1}^{(2)} & 0\\ 0 & {\sigma }_{2}^{(2)}\end{array}\right)\otimes {\bf{I}}\right)$$

where ⊗ is the Kronecker product and the combined vector of cell densities $(\mathbf{Y}_{i,k}, \mathbf{Y}_{j,k})$ of cell states i and j is modelled by a multivariate Gaussian distribution whose covariance decomposes into a spatial and a noise term. The spatial term

$$\left(\begin{array}{cc}{\sigma }_{1}^{(1)} & {\sigma }_{2}^{(1)}\\ {\sigma }_{2}^{(1)} & {\sigma }_{3}^{(1)}\end{array}\right)\otimes K\left({\bf{X}},l\right)$$

is defined by a between-cell-state covariance matrix

$$\left(\begin{array}{cc}{\sigma }_{1}^{(1)} & {\sigma }_{2}^{(1)}\\ {\sigma }_{2}^{(1)} & {\sigma }_{3}^{(1)}\end{array}\right)$$

and a spatial covariance matrix defined using the squared exponential kernel:

$$K{({\bf{X}},l)}_{mn}=\exp \left(-\frac{{\parallel {x}_{m}-{x}_{n}\parallel }^{2}}{2{l}^{2}}\right)$$

x_m and x_n are spatial coordinates of spots m and n and l is the length scale of the smooth Gaussian process function in space that is being fit to cell

densities.

The noise term

$$\left(\begin{array}{cc}{\sigma }_{1}^{(2)} & 0\\ 0 & {\sigma }_{2}^{(2)}\end{array}\right)\otimes {\bf{I}}$$

represents sources of variation other than spatial covariance of cell state densities.

The between-cell-state covariance matrix is constrained to be symmetric positive definite by defining

$$\left(\begin{array}{cc}{\sigma }_{1}^{(1)} & {\sigma }_{2}^{(1)}\\ {\sigma }_{2}^{(1)} & {\sigma }_{3}^{(1)}\end{array}\right)=\,\left(\begin{array}{cc}{a}_{1} & 0\\ {a}_{2} & {a}_{3}\end{array}\right){\left(\begin{array}{cc}{a}_{1} & 0\\ {a}_{2} & {a}_{3}\end{array}\right)}^{{\rm{T}}}$$

The free parameters a_1, a_2, a_3, $\sigma_1^{(2)}$, $\sigma_2^{(2)}$ and l are estimated using maximum likelihood and automatic differentiation in Tensorflow[60,61] using the BFGS algorithm. To improve convergence, we initialize l to the distance between centres of neighboring Visium spots.

This model allows us to infer which cell states are proximal in physical space and are likely to be migrating in the process of gradual differentiation in space.

For the spatial covariance model within StOrder workflow we only used a subset of our Visium data that corresponded to (1) decidua_and_villi_tips and (2) myometrium—because only these regions contained invading trophoblast cell states. For more details please see ‘Subsetting Visium data into anatomical regions with SpatialDE2’ in ‘Downstream analysis of 10x Genomics Visium data’ above. This helps to focus on the regions of the tissue that are relevant for the process of interest and is recommended to do in general if there are parts of the Visium data that do not contain cell states relevant to the process of interest.

Third, StOrder reconstructs connections between cell states by taking into account both the connectivity matrix (step 1) from single-cell transcriptomics data and the spatial covariance matrix (step 2) from the spatial data in the following way:

$${\beta }({\alpha }P+(1-{\alpha })S)+(1-{\beta })P\odot S$$

where P is the PAGA connectivity matrix, S is the spatial correlation matrix, α weights the contributions of P and S in the additive term, β weights the contributions of the additive and multiplicative terms, and $\odot$ is the element-wise product. It then reconstructs the putative trajectory tree using the built-in PAGA functions.

The combined connectivity matrix based on both gene expression and spatial data with a range of weight parameters revealed the fully resolved invasion trajectory tree of the EVT with the correct topology (all connected cell state components, one branching point, no cycles, start at VCT-CCC population and two end points: eEVT and GC populations). The choice of ω parameter (contribution/weight of gene expression vs spatial part in the final matrix) in this last step depends on the goal of using this approach. In our case, we assumed: (1) the origin of EVT (VCT-CCC); (2) the end points of EVT (eEVT and GC); (3) the determination of a single branching point; and (4) the absence of cyclic trajectory. We therefore produced trajectory trees for 10,201 of (α,β) value pairs (from 0 to 1 with 0.01 increment step each) representative of different tree topologies corresponding to different ratios of gene expression vs spatial contribution. Out of the 10,201 tree structures we inspected, 3,574 trees represented the topology with the assumptions described above. These trajectories consistently assigned EVT-2 as the putative branching point. Tree structures with mainly gene expression-based connectivity values did not yield a branching point population we were looking for. Tree structures with mainly spatial based connectivities hindered the link between iEVT and GC populations, likely due to the large length scale of this invasion in space.

Limitations

Our approach assumes the gradual nature of gene expression changes accompanied by gradual migration of cells in space while they differentiate. Thus, it may not yield meaningful results in scenarios where this underlying assumption is violated. In addition, it is recommended that the user estimates the spatial scale at which the process of interest is taking place—whether in current Visium resolution the differentiation and migration is happening over the course of only a few spots or many more—this will change the initial values of l parameter and help the model fit the data better.

Combined RNA and ATAC analysis using MEFISTO

Preprocessing of multiome data and training of the MEFISTO model

Gene expression (snRNA-seq) counts of the multiome data for donor P13 were normalized by total counts (scanpy.pp.normalize_per_cell(rna, counts_per_cell_after=1e4)) and log-transformed (pp.log1p(rna)). Highly variable gene features were then calculated (sc.pp.highly_variable_genes(rna, min_mean=0.0125, max_mean=3, min_disp=0.5)) and the subsetted object's expression was scaled (sc.pp.scale(rna, max_value=10)).

Chromatin accessibility (scATAC-seq) counts of the multiome data for donor P13 were preprocessed using TF-IDF normalization (muon.atac.pp.tfidf(atac[key], scale_factor=1e4)). To select biologically meaningful highly variable peak features, ATAC counts were aggregated into pseodubulks by cell states and averaged, then variance of accessibility was calculated across these pseudobulks, and informative peak features were selected based on this measure (top 75th percentile (10,640) of peaks selected in total) as the peaks with highest variance. Finally, these data were scaled (sc.pp.scale(atac, max_value=10)).

Using the preprocessed RNA and ATAC data we used a pseudotime-aware dimensionality reduction method MEFISTO[30] to extract major sources of variation from the RNA and ATAC data jointly and identify coordinated patterns along the invasion trajectory. As a proxy for the trophoblast invasion trajectory in the MEFISTO model we used 2-dimensional pseudotime coordinates based on a UMAP of the RNA data by calculating

PCA (sc.tl.pca(rna, n_comps=8)), neighborhood graph (sc.pp.neighbors(rna)) and UMAP embedding (sc.tl.umap(rna)).

The MEFISTO model was trained using the following command within MUON (version 0.1.2) package interface:

muon.tl.mofa(mdata, outfile='',

use_obs = "union",

smooth_covariate=["UMAP1", "UMAP2"],

use_float32=True)

We further excluded factor 5 from downstream analysis as a technical artefact due to its significant and high correlation (Spearman rank-order correlation coefficient 0.94 (over all cell states), $P < 10^{-308}$, two-sided test) with the n_peaks_by_counts (number of ATAC peaks with at least 1 count in a nucleus) in ATAC view in all cell states (Supplementary Fig. 4k) and lack of smoothness along pseudotime (Supplementary Fig. 4j).

Defining groups of ATAC peak features

To further interpret ATAC features, we annotated them based on their genomic location using GenomicRanges package (version 1.42.0). In parallel, we used epigenetic data from[62] to mark peak features in close proximity to trophoblast-specific enhancer features. To do so, we used peak files corresponding to H3K4me1, H3K27ac and H3K27me3 histone modifications marks for second trimester trophoblast samples (obtained from authors of aforementioned study upon request) to infer regions of the genome corresponding to active (H3K27ac + H3K27me3), primed (only H3K4me1) or repressed (H3K4me1 + H3K27me3) enhancers. This was done using bedtools (version 2.30.0) in the following way:

1. (1)

   bedtools subtract -a H3K4me1_file.bed -b H3K27ac_file.bed > interm_file.bed bedtools subtract -a interm_file.bed -b H3K27me3_file.bed > primed_enhancers.bed To produce primed enhancers file

2. (2)

   bedtools intersect -a H3K4me1_file.bed -b H3K27ac_file.bed > active_enhancers.bed To produce active enhancers file

3. (3)

   bedtools intersect -a H3K4me1_file.bed -b H3K27me3_file.bed > repressed_enhancers.bed To produce repressed enhancers file

The enhancer files produced were then overlapped with peaks in ATAC analysis (bedtools intersect -a atac_peaks_file.bed -b enhancer_file.bed -wa) and any peaks having a >1-bp overlap with an enhancer feature were considered to be proximal to those features (done separately for active, primed and repressed enhancers).

Enrichment analysis of features in the MEFISTO model

Gene set enrichment analysis for gene features was performed based on the C5 category and the Biological Process subcategory from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb) using GSEA functionality implemented in MOFA2 (run_enrichment command, MOFA2 version 1.3.5). This was done separately for negative and positive weights of each factor.

Peak group enrichment for peak features was performed using the same run_enrichment command in MOFA2 on peak groups defined as described above (Defining groups of ATAC peak features).

Transcription factor analysis using the MEFISTO model

To extract information about transcription factor binding motif enrichment in ATAC features of MEFISTO factors, we first performed enrichment analysis of peaks using GSEA functionality implemented in MOFA2 (run_enrichment command, MOFA2 version 1.3.5) on the peak-motif matrix produced by Signac package (version 1.5.0). Then, to identify which MEFISTO factors contribute the most to each transition of cell states along the invading trophoblast trajectory (inferred with StOrder), we trained logistic regression classifiers for each transition along the trajectory (overall for 6 transitions: VCT→VCT-CCC, VCT-CCC→EVT-1, EVT-1→EVT-2, EVT-2→iEVT, iEVT→GC, EVT-2→eEVT) on the matrix of factor values. For each transition the factor with the highest absolute coefficient separating the two cell states was selected, accounting for the sign of contribution in the logistic regression (positive or negative). If the top factor is contributing to a transition with a positive coefficient, transcription factor binding motifs coming from MEFISTO enrichment analysis of this factor's top positive values are further considered in general transcription factor analysis as transcription factors upregulated upon this transition, whereas transcription factor binding motifs coming from MEFISTO enrichment analysis of this factor's top negative values are further considered in general transcription factor analysis as transcription factors downregulated upon this transition. All of these transcription factor motifs are marked as having evidence from the MEFISTO factor relevant for this transition. Reverse

procedure is applied in case if the top factor is contributing to a transition with a negative coefficient in the corresponding logistic regression model.

For more details please see the following notebook: https://github.com/ventolab/MFI/blob/main/2_inv_troph_trajectory_and_TFs/2-5_MEFISTO_analysis_inv_troph/S3_DEG_comparison_to_MEFISTO_factor_translation.ipynb

Trophoblast trajectory inference analysis

To derive trophoblast pseudotime based on transcriptomic similarity, we used Slingshot v1.8.0. With Slingshot we fitted a cluster-based minimum spanning tree (MST) over the two-dimensional UMAP of P13 trophoblasts, and inferred the global lineage topology to assign cell states to lineages. Only donor P13 cells in the G1 phase of the cell cycle were included. To balance trophoblast state contributions, we downsampled each trophoblast state to account for up to 100 cells per state. VCT was assigned as the initial cell state (start.clus), while eEVT, SCT and GC were assigned as terminal states (end.clus). Slingshot fits simultaneous principle curves to smooth the MST and assigns a weight for each trophoblast cell in each lineage. Slingshot outputs lineage-specific pseudotimes and weights of assignment for each cell.

We next fitted a tradeSeq (v1.4.0) gene expression model (negative binomial generalized additive model) using the trajectory pseudotime and the weights computed with Slingshot (with nknots = 6). Next, we tested whether the gene expression is significantly changing along trophoblast pseudotime. For such a purpose, we used the statistical test implemented in the associationTest function, which tests the null hypothesis that all smoother coefficients are equal to each other. Genes with a $P < 10^{-6}$ and mean logFC > 0.5 were selected as the main drivers of the trophoblast trajectory.

Cell label transferring on trophoblast organoids

To transfer cell labels from donor P13 snRNA-seq in vivo trophoblast to the scRNA-seq TSC and PTO we trained two independent logistic regression models. The P13 dataset was downsampled to 500 cells per trophoblast state, except for GC and eEVT, which were discarded from the training due to their scarcely abundance. The common highly variable genes (1,695 genes for PTO and 1,565 for TSC), of the 4,000 selected per dataset, between the in vivo and each individual organoid dataset were selected as features for model training. The in vivo dataset was split into 80/20 training and test set, hyperparameters were explored employing a threefold cross-validation and scored employing the mean Matthews correlation coefficient of each fold. Top-ranked models were selected and assessed on the test set, with no significant differences found between them. Finally, the best model for each organoid dataset was employed to transfer cell labels from donor P13.

Cell–cell communication analysis with CellPhoneDB

To retrieve interactions between invading trophoblast and other cell populations identified in our samples, we used the CellPhoneDB degs_analysis method[13,63] (https://github.com/ventolab/CellphoneDB) described in ref. [33]. In short, we retrieved the interacting pairs of ligands and receptors meeting the following requirements: (1) all the protein members were expressed in at least 10% of the cell type under consideration; and (2) at least one of the protein members in the ligand or the receptor was a DEG in an invading trophoblast subset (according to our analysis of differential expression, for details please see 'Differential gene expression analysis'), with an adjusted *P*-value below 0.05 and $\log FC > 0.1$. We further selected which cell states are spatially co-located in each microenvironment via visual inspection of cell2location deconvolution results for our Visium data. The analysis was done on an updated version of CellPhoneDB-database (v4.1) which includes novel intercellular interactions from refs. [64,65]. Only bona fide manually curated interactions were considered in the analysis.

Transcription factor analysis

To prioritize the transcription factors relevant for each invading trophoblast cell state or microenvironment, we integrate four types of measurements: (1) expression levels of the transcription factor and (2) the activity status of the transcription factor measured from (2a) the expression levels of their targets (described in 'Transcription factor activities derived from scRNA-seq and snRNA-seq') and/or (2b) the chromatin accessibility of their binding motifs (described in 'Transcription factor motif activity analysis from scATAC–seq') and/or (2c) evidence of the chromatin accessibility of their binding motifs in relevant factors from multimodal RNA-ATAC analysis (with MEFISTO). Plots in main figures include transcription factor meeting the following criteria: (1) transcription factor was differentially expressed, with adjusted P-value < 0.05) and/or (2) transcription factor was differentially active, with $\log_2$ fold change greater than 0.25 and adjusted P-value < 0.05 in at least one of the transcription factor activity measurements (2a or 2b).

Transcription factor differential expression from scRNA-seq and snRNA-seq

We compute differential expression using the procedure described in 'Differential gene expression analysis' and further subset resulting gene targets to transcription factors only based on the list of transcription factors provided by DoRothEA.

Transcription factor activities derived from scRNA-seq and snRNA-seq

We estimated protein-level activity for human transcription factor as a proxy of the combined expression levels of their targets. Target genes were retrieved from Dorothea[66], an orthogonal collection of transcription factor targets compiled from a range of different sources. Next, we estimated transcription factor activities for each cell using Viper[67], a GSEA-like approach, as implemented in the Dorothea R package and tutorial[68] for the genes differentially expressed along the invading trophoblast trajectory (see 'Differential gene expression analysis').

Transcription factor motif activity analysis from scATAC–seq

Transcription factor motif activities were computed using chromVar[69] v. 1.12.2 with positional weight matrices from JASPAR2018[70], HOCOMOCOv10[71], SwissRegulon[72], HOMER[73]. chromVar returns a matrix with binding activity estimates of each transcription factor in each cell, which we used to test for differential transcription factor binding activity between trophoblast cell states with FindMarkers function in Seurat (default parameters) in the same way as described in 'Differential gene expression analysis' (backwards along invading trophoblast trajectory).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Open access datasets are available from ArrayExpress (www.ebi.ac.uk/arrayexpress), with accession numbers E-MTAB-12421 (scRNA-seq and snRNA-seq of primary tissue), E-MTAB-12595 (multiome snRNA-seq and snATAC-seq), E-MTAB-12698 (Visium), E-MTAB-12650 (scRNA-seq and snRNA-seq of PTOs). Managed-access datasets are available from EGA archive (https://ega-archive.org/) with accession numbers EGAD00001010037 (scRNA-seq and snRNA-seq of historical placental beds), EGAD00001010038 (multiome snRNA-seq and snATAC-seq of historical placental beds), EGAD00001010017 (scRNA-seq and snRNA-seq of TSCs). Image datasets are available at the EMBL–EBI BioImage Archive (https://www.ebi.ac.uk/biostudies) under accession number S-BIAD615. All datasets are public access. scRNA-seq and snRNA-seq datasets to reproduce UMAPs and dot plots can be accessed and downloaded through the web portals at https://www.reproductivecellatlas.org. The external scRNA-seq dataset of the first-trimester human decidual–placental interface is available from ArrayExpress (E-MTAB-6701).

Code availability

All code used for data analysis is available at https://github.com/ventolab/MFI[74].

References

1. Pijnenborg, R., Vercruysse, L. & Hanssens, M. The uterine spiral arteries in human pregnancy: facts and controversies. *Placenta* **27**, 939–958 (2006).

2. Brosens, I., Pijnenborg, R., Vercruysse, L. & Romero, R. The 'great obstetrical syndromes' are associated with disorders of deep placentation. *Am. J. Obstet. Gynecol.* **204**, 193–201 (2011).

3. Turco, M. Y. et al. Trophoblast organoids as a model for maternal–fetal interactions during human placentation. *Nature* **564**, 263–267 (2018).

4. Haider, S. et al. Self-renewing trophoblast organoids recapitulate the developmental program of the early human placenta. *Stem Cell Rep.* **11**, 537–551 (2018).

5. Okae, H. et al. Derivation of human trophoblast stem cells. *Cell Stem Cell* **22**, 50–63.e6 (2018).

6. Turco, M. Y. & Moffett, A. Development of the human placenta. *Development* **146**, dev163428 (2019).

7. Burton, G. J. & Jauniaux, E. The cytotrophoblastic shell and complications of pregnancy. *Placenta* **60**, 134–139 (2017).

8. Jauniaux, E. et al. Onset of maternal arterial blood flow and placental oxidative stress. A possible factor in human early pregnancy failure. *Am. J. Pathol.* **157**, 2111–2122 (2000).

9. Burton, G. J., Woods, A. W., Jauniaux, E. & Kingdom, J. C. P. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. *Placenta* **30**, 473–482 (2009).

10. al-Lamki, R. S., Skepper, J. N. & Burton, G. J. Are human placental bed giant cells merely aggregates of small mononuclear trophoblast cells? An ultrastructural and immunocytochemical study. *Hum. Reprod.* **14**, 496–504 (1999).

11. Garrido-Gomez, T. et al. Defective decidualization during and after severe preeclampsia reveals a possible maternal contribution to the etiology. *Proc. Natl Acad. Sci. USA* **114**, E8468–E8477 (2017).

12. Jauniaux, E., Jurkovic, D., Hussein, A. M. & Burton, G. J. New insights into the etiopathology of placenta accreta spectrum. *Am. J. Obstet. Gynecol.* **227**, 384–391 (2022).

13. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018).

14. Sheridan, M. A. et al. Characterization of primary models of human trophoblast. *Development* **148**, dev199749 (2021).

15. Karvas, R. M. et al. Stem-cell-derived trophoblast organoids model human placental development and susceptibility to emerging pathogens. *Cell Stem Cell* **29**, 810–825.e8 (2022).

16. Shannon, M. J. et al. Single-cell assessment of trophoblast stem cell-based organoids as human placenta-modeling platforms. Preprint at *bioRxiv* https://doi.org/10.1101/2022.11.02.514970 (2022).

17. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).

18. Shannon, M. J. et al. Cell trajectory modeling identifies a primitive trophoblast state defined by BCAM enrichment. *Development* **149**, dev199840 (2022).

19. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).

20. Mi, S. et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).

21. Suryawanshi, H. et al. A single-cell survey of the human first-trimester placenta and decidua. *Sci. Adv.* **4**, eaau4788 (2018).

22. Shen, L. et al. Lysophosphatidylcholine acyltransferase 1 promotes epithelial-mesenchymal transition of hepatocellular carcinoma via the Wnt/β-catenin signaling pathway. *Ann. Hepatol.* **27**, 100680 (2022).

23. Haider, S. et al. Notch1 controls development of the extravillous trophoblast lineage in the human placenta. *Proc. Natl Acad. Sci. USA* **113**, E7710–E7719 (2016).

24. Lee, C. Q. E. et al. Integrin α2 marks a niche of trophoblast progenitor cells in first trimester human placenta. *Development* **145**, dev162305 (2018).

25. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

26. Burrows, T. D., King, A. & Loke, Y. W. Expression of adhesion molecules by endovascular trophoblast and decidual endothelial cells: Implications for vascular invasion during implantation. *Placenta* **15**, 21–33 (1994).

27. Kam, E. P., Gardner, L., Loke, Y. W. & King, A. The role of trophoblast in the physiological change in decidual spiral arteries. *Hum. Reprod.* **14**, 2131–2138 (1999).

28. Chang, W.-L. et al. PLAC8, a new marker for human interstitial extravillous trophoblast cells, promotes their invasion and migration. *Development* **145**, dev148932 (2018).

29. Gaccioli, F., Aye, I. L. M. H., Sovio, U., Charnock-Jones, D. S. & Smith, G. C. S. Screening for fetal growth restriction using fetal biometry combined with maternal biomarkers. *Am. J. Obstet. Gynecol.* **218**, S725–S737 (2018).

30. Velten, B. et al. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. *Nat. Methods* **19**, 179–186 (2022).

31. Nguyen, B.-C. et al. Cross-regulation between Notch and p63 in keratinocyte commitment to differentiation. *Genes Dev.* **20**, 1028–1042 (2006).

32. Moffett, A. & Shreeve, N. Local immune recognition of trophoblast in early human pregnancy: controversies and questions. *Nat. Rev. Immunol.* https://doi.org/10.1038/s41577-022-00777-2 (2022).

33. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).

34. Varberg, K. M. et al. ASCL2 reciprocally controls key trophoblast lineage decisions during hemochorial placenta development. *Proc. Natl Acad. Sci. USA* **118**, e2016517118 (2021).

35. Kong, S.-Y. et al. The ELK3–GATA3 axis orchestrates invasion and metastasis of breast cancer cells in vitro and in vivo. *Oncotarget* **7**, 65137–65146 (2016).

36. Froese, N. et al. GATA6 promotes angiogenic function and survival in endothelial cells by suppression of autocrine transforming growth factor beta/activin receptor-like kinase 5 signaling. *J. Biol. Chem.* **286**, 5680–5690 (2011).

37. Garcia-Alonso, L. et al. Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).

38. Jokhi, P. P., Chumbley, G., King, A., Gardner, L. & Loke, Y. W. Expression of the colony stimulating factor-1 receptor (c-fms product) by cells at the human uteroplacental interface. *Lab. Invest.* **68**, 308–320 (1993).

39. Matloubian, M., David, A., Engel, S., Ryan, J. E. & Cyster, J. G. A transmembrane CXC chemokine is a ligand for HIV-coreceptor Bonzo. *Nat. Immunol.* **1**, 298–304 (2000).

40. Bojić-Trbojević, Ž. et al. Human trophoblast requires galectin-3 for cell migration and invasion. *Sci. Rep.* **9**, 2136 (2019).

41. Pijnenborg, R., Dixon, G., Robertson, W. B. & Brosens, I. Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta* **1**, 3–19 (1980).

42. Brosens, I., Robertson, W. B. & Dixon, H. G. The physiological response of the vessels of the placental bed to normal pregnancy. *J. Pathol. Bacteriol.* **93**, 569–579 (1967).

43. Burton, G. J., Jauniaux, E. & Watson, A. L. Maternal arterial connections to the placental intervillous space during the first trimester of human pregnancy: the Boyd collection revisited. *Am. J. Obstet. Gynecol.* **181**, 718–724 (1999).

44. King, A. & Loke, Y. W. Differential expression of blood-group-related carbohydrate antigens by trophoblast subpopulations. *Placenta* **9**, 513–521 (1988).

45. Smith, G. C. S. First-trimester determination of complications of late pregnancy. *JAMA* **303**, 561–562 (2010).

46. Kenny Roberts, L. T. Embedding and freezing fresh human tissue in OCT using isopentane V.3. *protocols.io* https://doi.org/10.17504/protocols.io.95mh846 (2019).

47. Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).

48. Sheridan, M. A. et al. Establishment and differentiation of long-term trophoblast organoid cultures from the human placenta. *Nat. Protoc.* **15**, 3441–3463 (2020).

49. Bayraktar, O. A. et al. Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat. Neurosci.* **23**, 500–509 (2020).

50. Popescu, D.-M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).

51. Heaton, H. et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020).

52. Park, J.-E. et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).

53. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324.e18 (2018).

54. Gaspar, J. M. Improved peak-calling with MACS2. Preprint at *bioRxiv* https://doi.org/10.1101/496521 (2018).

55. Kleshchevnikov, V. et al. Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *Nat. Biotechnol.***40**, 661–671 (2022).

56. Kats, I., Vento-Tormo, R. & Stegle, O. SpatialDE2: fast and localized variance component analysis of spatial transcriptomics. Preprint at *biorXiv* https://doi.org/10.1101/2021.10.27.466045 (2021).

57. González-Blas, C. B. et al. cisTopic: *cis*-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).

58. Bravo González-Blas, C. et al. Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol. Syst. Biol.* **16**, e9438 (2020).

59. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).

60. Nocedal, J. & Stephen J. Wright, S. J. Numerical Optimization. in *Springer Series in Operations Research and Financial Engineering* (Springer, 2000).

61. TensorFlow Developers. TensorFlow https://doi.org/10.5281/zenodo.6519082 (2022).

62. Zhang, B. et al. Human placental cytotrophoblast epigenome dynamics over gestation and alterations in placental disease. *Dev. Cell* **56**, 1238–1252.e5 (2021).

63. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).

64. Shilts, J. et al. A physical wiring diagram for the human immune system. *Nature* **608**, 397–404 (2022).

65. Kanemaru, K. *et al*. Spatially resolved multiomics of human cardiac niches. Preprint at *bioRxiv* https://doi.org/10.1101/2023.01.30.526202 (2023).

66. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).

67. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).

68. Holland, C. H. et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).

69. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from

single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

70. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

71. Kulakovskiy, I. V. et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res.* **44**, D116–D125 (2016).

72. Pachkov, M., Erb, I., Molina, N. & van Nimwegen, E. SwissRegulon: a database of genome-wide annotations of regulatory sites. *Nucleic Acid Res.* **35**, D127–D131 (2007).

73. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

74. Arutyunyan, A. *ventolab/MFI: Publishing MFI analysis code* https://doi.org/10.5281/zenodo.7565015 (2023).

Acknowledgements

This publication is part of the Human Cell Atlas. The authors thank the Sanger Cellular Generation and Phenotyping (CGaP) Core Facility and the Sanger Core Sequencing pipeline for support with sample processing and sequencing library preparation; A. Surani for supplying the TSC lines; H. Okae and T. Arima for sharing permission; R. Argelaguet, V. Kleshchevnikov, S. van Dongen, M. Prete and S. Murray for insightful comments and web portal support; T. Porter and the Cellular Genetics wet lab team for experimental support; A. Garcia for graphical images; and A. Maartens for editing. Placental material was provided by the Joint MRC–Human Cell Atlas (MR/S036350/1). The authors are grateful to patients for donating tissue for research. We thank D. Moore and M. Maquinana and staff at Addenbrooke's Hospital, Cambridge, UK. Supported by Wellcome Sanger core funding (WT206194 and 220540/Z/20/A) and the Wellcome Trust grant 'Wellcome Strategic Support Science award' (grant no.

211276/Z/18/Z). M.Y.T. held the Royal Society Dorothy Hodgkin Fellowship (DH160216) and Research Grant for Research Fellows (RGF\R1\180028) during this study and is also supported by funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (Grant agreement 853546). A.M. is in receipt of a Wellcome Trust Investigator Award (200841/Z/16/Z).

Author information

Author notes

1. Margherita Y. Turco

   Present address: Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland

2. These authors contributed equally: Anna Arutyunyan, Kenny Roberts, Kevin Troulé, Frederick C. K. Wong, Megan A Sheridan

Authors and Affiliations

1. Wellcome Sanger Institute, Cambridge, UK

   Anna Arutyunyan, Kenny Roberts, Kevin Troulé, Frederick C. K. Wong, Luz Garcia-Alonso, Britta Velten, Regina Hoo, Elias R. Ruiz-Morales, Carmen Sancho-Serra, Jarrod Shilts, Louis-Francois Handfield, Elizabeth Tuck, Cecilia Icoresi Mazzeo, Iva Kelava, Elena Prigmore, Sarah A. Teichmann, Omer Ali Bayraktar, Oliver Stegle & Roser Vento-Tormo

2. Centre for Trophoblast Research, University of Cambridge, Cambridge, UK

   Anna Arutyunyan, Megan A. Sheridan, Lucy Gardner, Qian Li, Ashley Moffett, Margherita Y. Turco & Roser Vento-Tormo

3. Department of Pathology, University of Cambridge, Cambridge, UK

Megan A. Sheridan, Lucy Gardner, Qian Li, Ashley Moffett & Margherita Y. Turco

4. Division of Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany

   Ilia Kats, Britta Velten, Luca Marconato & Oliver Stegle

5. European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany

   Luca Marconato & Oliver Stegle

6. Department of Biology, Hull York Medical School, York Biomedical Research Institute, University of York, York, UK

   Gavin J. Wright

7. Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, UK

   Sarah A. Teichmann

Contributions

R.V.-T. conceived and designed the study with contributions from A.A., A.M. and M.Y.T. A.M. collected and analysed the in vivo historical samples. E.P. and R.H. performed the nuclei experiments. K.R. and E.T. performed the spatial transcriptomics analyses with help of C.I.M. and I.Kelava Q.L. and L.G. performed immunohistochemistry experiments. F.C.K.W. performed TSC experiments with help from E.R.R.-M. and C.S.-S., J.S. and G.J.W. M.A.S. performed PTO experiments. A.A. and K.T. analysed the data with help from B.V., L.G.-A., L.F.-H. and L.M. A.A. and I.Kats developed StOrder. A.A., K.T. and R.V.-T. interpreted the data with contribution of M.Y.T., A.M., M.A.S., S.A.T. and K.R. R.V.-T. supervised the work with help from O.S. (StOrder), M.Y.T. (in vitro PTO benchmark) and O.A.B. (spatial transcriptomics). R.V.-T. and A.M. wrote the

manuscript with contributions from K.R., M.A.S., and A.A. The final version of the manuscript has been approved by all the authors.

Corresponding authors

Correspondence to Omer Ali Bayraktar, Ashley Moffett, Oliver Stegle, Margherita Y. Turco or Roser Vento-Tormo.

Ethics declarations

Competing interests

S.A.T. has received remuneration for consulting and scientific advisory board work from Genentech, Biogen, Roche and GlaxoSmithKline as well as Foresite Labs over the past three years. O.S. is a paid scientific advisory board member of Insitro Inc.

Peer review

Peer review information

Nature thanks John Aplin, Dominic Gruen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Spatial transcriptomics of human placental bed.

a: High-resolution imaging of a section of the placenta-decidua interface stained by in situ hybridization (smFISH) for *HLA-G*, illustrating the depth of invasion of EVTs into the uterus (n = 1). Magnified insets (dashed squares) highlight the HLA-G-negative placental villi, and *HLA-G*+ EVTs emerging from the CCC to invade the decidua and myometrium. **b**: Overview of experimental design of the study. **c**: Cohort composition split by gestational age window (post-conceptional weeks, PCW) representing tissues sampled from each donor and performed assays. Highlighted in red rectangles are the three donors whose tissues have been additionally profiled with spatial transcriptomics (Visium) and multiome assays. **d**: Histological overview (H&E staining) of donors P13, P14 and Hrv43 tissues with annotations of tissue regions. For the implantation site of donor P13 (~ 8-9 PCW, left); black squares (small) indicate trophoblast microenvironments in space; faint grey squares (large) indicate positioning of tissue on Visium spatial transcriptomics capture areas; arrow indicates representative Visium section further explored in Fig. 1d. For Visium, P13 (n = 5 feature areas, 4 consecutive slides with overlapping positions and 1 slide from an additional tissue block - P13b), P14 (n = 2 feature areas, consecutive slides with same position), Hrv43 (n = 1 feature area). **e**: Cell state locations (derived with cell2location) in representative Visium sections of donors P14 and Hrv43 highlighting relevant spatial trophoblast microenvironments. Spot colour indicates cell state densities computed by cell2location as the number of cells of a given cell state in a Visium spot. Cytotrophoblast cell column (CCC), extravillous trophoblast (EVT), interstitial EVT (iEVT), giant cells (GC), endovascular EVT (eEVT), single-cell RNA sequencing (scRNA-seq), single-nuclei RNA sequencing (snRNA-seq), microenvironment (ME), Hematoxylin and Eosin (H&E).

Extended Data Fig. 2 scRNA-seq and snRNA-seq data quality control and data analysis overview.

a: Overview of the computational pipeline implemented for analysis of *in vivo* scRNA-seq and snRNA-seq data. Data integrated with scVI. **b**–**e**: (top) UMAP (uniform manifold approximation and projection) scatterplots of donors P13 (n = 67,821 nuclei), P14 (n = 45,166 nuclei), Hrv43 (n = 60,837 nuclei and cells) and all donors' (n = 325,665 nuclei and cells, m = 18 donors) scRNA-seq and snRNA-seq data (b-e respectively) for all

recovered cell states, coloured by coarse grain compartment annotation and metadata labels: assay, sample (10X library), donor and developmental age. (bottom) Dot plots show normalised, log-transformed and variance-scaled expression of genes characteristic of coarse grain compartment (X-axis) in donors profiled (Y-axis). Single-cell RNA sequencing (scRNA-seq), single-nuclei RNA sequencing (snRNA-seq), maternal (m), fetal (f), natural killer (NK), innate lymphocytes (ILC), single-cell RNA sequencing (scRNA-seq), single-nuclei RNA sequencing (snRNA-seq).

Extended Data Fig. 3 snRNA-seq and scRNA-seq trophoblast data analysis overview.

a: UMAP (uniform manifold approximation and projection) scatterplots of donor P13 snRNA-seq data (n = 37,675 nuclei) for all trophoblast cell states coloured by (from left to right) assay, cell cycle phase of the nuclei and sample (10X library). Please note: bioinformatics analyses used cannot distinguish between G0 and G1. **b**: UMAP scatterplot of integrated snRNA-seq and scRNA-seq of all donors' (n = 75,042 nuclei and cells, m = 17 donors with trophoblast present) trophoblast cell states in the maternal-fetal interface coloured by cell state. **c**: UMAP scatterplots of all donors' scRNA-seq and snRNA-seq data for all donors' (n = 75,042 nuclei and cells, m = 17 donors with trophoblast present) trophoblast cell states coloured by assay, sample (10X library), cell cycle phase of the cells/nuclei, donor and developmental age. **d**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (Y-axis) characteristic of trophoblast cell states (X-axis) in all donors (m = 17 donors with trophoblast present). **e**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) characteristic of VCT cell states (Y-axis) in all donors (m = 17 donors with trophoblast present). **f**: Results of PAGA trajectory inference of all trophoblast cell states in donor P13 snRNA-seq data. (left) main manifold, center: denoised PAGA manifold, (right) PAGA reconstruction of putative trajectory tree for all trophoblast cell states. For the purpose of this analysis all EVTs have been united in annotation under the 'EVT' label. Syncytiotrophoblast (SCT), villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), proliferative (p), extravillous trophoblast (EVT), interstitial EVTs (iEVTs),

giant cells (GC), endovascular EVT (eEVT), single-cell RNA sequencing (scRNA-seq), single-nuclei RNA sequencing (snRNA-seq).

Extended Data Fig. 4 Analysis of extravillous trophoblast invasion trajectory using stOrder.

a: Schematic overview of StOrder approach representing the workflow of joint cell differentiation trajectory inference from gene expression and spatial transcriptomics data (See methods). **b**: (Left) Main UMAP (uniform manifold approximation and projection) scatterplot and (right) denoised manifold used for PAGA trajectory inference of all trophoblast cell states in donor P13 single-nuclei RNA sequencing (snRNA-seq) data. **c**: PAGA reconstruction of putative trajectory tree for all extravillous trophoblast cell states. This corresponds to the trajectory tree inferred by stOrser with $\alpha = 1$, $\beta = 1$ parameters from donor P13 snRNA-seq data and spatial transcriptomics data of donors P13 (5 capture areas), P14 (2 capture areas) and Hrv43 (1 capture area). **d**: Heatmap of binary success matrix of stOrder approach for pairs of (α,β), values (along Y and X axes, respectively). Assigned matrix value is 1 if a tree of correct topology has been reconstructed for that pair of (α,β) values, and 0 if no tree of correct topology was reconstructed. **e**: Reconstruction of putative invading trophoblast trajectory tree based solely on spatial data inferred by stOrder with $\alpha = 0$, $\beta = 1$ parameters from donor P13 snRNA-seq data and spatial transcriptomics data of donors P13 (5 chips), P14 (2 chips) and Hrv43 (1 chip). Villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), extravillous trophoblast (EVT), interstitial EVTs (iEVTs), giant cells (GC), endovascular EVT (eEVT).

Extended Data Fig. 5 Analysis of extravillous trophoblast invasion trajectory using Slingshot.

a: Minimum spanning tree of donor P13 trophoblast single-nuclei RNA sequencing (snRNA-seq) data computed by Slingshot, visualised on the UMAP (uniform manifold approximation and projection) embedding of P13 donor trophoblast cells from Fig. 1c (n = 37,675 nuclei). Bigger black dots indicate trophoblast states. Smaller dots' colour indicates pseudotime.

b: Heatmap showing normalised and log-transformed expression values of the 567 genes associated with trophoblast pseudotime, estimated with tradeSeq (p-value < 1x10-6 and mean LogFC > 0.5). Cells in rows are sorted according to the predicted pseudotime. Genes are sorted according to the trophoblast state where gene expression peaks. Marker genes are highlighted. Syncytiotrophoblast (SCT), villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), extravillous trophoblast (EVT), interstitial EVTs (iEVTs), giant cells (GC), endovascular EVT (eEVT).

Extended Data Fig. 6 Histological characterisation of eEVTs.

a: Estimated amount of mRNA computed by cell2location (colour intensity) contributed by each cell population to each spot (colour) shown over the hematoxylin and eosin (H&E) image of donor P13 implantation site (n = 1). **b**: (Top) High-resolution imaging of sections of the placenta-decidua interface from two donors (n = 2, donor ID is indicated in each panel), stained by multiplexed single molecule fluorescence in situ hybridization (smFISH) for *HLA-G* and *NCAM1*; dashed squares indicate areas shown magnified below. (Middle) magnified insets highlight an artery containing aggregating eEVTs (left) and cytotrophoblast cell columns; in the latter, solid arrows indicate sporadic nascent *NCAM1*+ cells shown magnified below (bottom). **c**: Expression of NCAM1 (marker of eEVT) with IHC in first-trimester decidual tissue. Nuclei are counterstained with hematoxylin. Representative images from three different donors (n = 3).

Extended Data Fig. 7 Multimodal analysis of extravillous trophoblast invasion.

a: Overview of the computational pipeline implemented for analysis of multimodal data. **b**-**c**: UMAP (uniform manifold approximation and projection) scatterplot of multimodal snATAC-seq data from donors P13, P14 and Hrv43 (n = 52,798 nuclei) coloured by cell state (b), donor, sample (10X library) or unbiased clustering labels (c). Data is annotated based on the corresponding single-nuclei RNA sequencing (snRNA-seq) cell state assignment. **d**: UMAP scatterplot of integrated multimodal single-nuclei ATAC sequencing (snATAC-seq) data for trophoblast only from donors P13, P14 and Hrv43 (n = 7449 nuclei) coloured by cell state, donor and

sample (10X library). **e**: UMAP scatterplot of multiome (snRNA-snATACseq) data of invading trophoblast cells from donor P13 (n = 1605 nuclei) coloured by cell state. The manifold is calculated based on dimensionality reduction performed by MEFISTO (model with n = 9 factors). **f**: (Left) UMAP scatterplot of multiome (snRNA-snATACseq) data of invading trophoblast cells from donor P13 (n = 1605 nuclei) coloured by sample. The manifold is calculated based on dimensionality reduction performed by MEFISTO (model with n = 9 factors). (Right) Scatterplot of UMAP coordinates obtained from the RNA expression data that were used as covariates for MEFISTO, coloured by cell state. **g**: Heatmap representing percentage of variance explained by each MEFISTO factor in each data modality. **h**: Smoothness along differentiation estimated with MEFISTO. **i**: UMAP scatterplot of multiome (snRNA-snATACseq) data of invading trophoblast cells from donor P13 (n = 1605 nuclei) coloured by cell cycle phase and MEFISTO factor values for important selected factors. **j**: Spearman's rank correlation coefficients of each latent factor learned with MEFISTO and the number of genes per counts in snATAC-seq data (multiome). **k**: Gene set (RNA) enrichment analysis overview of MEFISTO factor 2 using two-sided parametric t-test, FDR is used to rank gene sets. **l**: Peak set (ATAC) enrichment analysis overview of MEFISTO factor 10 using two-sided parametric t-test, FDR is used to rank peak sets. Villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), proliferative (p), extravillous trophoblast (EVT), interstitial EVTs (iEVTs), giant cells (GC), endovascular EVT (eEVT), dendritic cells (DC), lymphatic (l), maternal (m), fetal (f) Hofbauer cells (HOFB), innate lymphocytes (ILC), macrophages (M), monocytes (MO), natural killer (NK), perivascular (PV), decidual (d), epithelial (epi), stromal (S), fibroblasts (F), uterine smooth muscle cells (uSMC).

Extended Data Fig. 8 Transcription factors active in extravillous trophoblast cell states (all donors).

a: Heatmap showing z-score of normalised, log-transformed and variance-scaled expression of transcription factors (TFs) upregulated during trophoblast invasion in all donors (n = 17 donors with trophoblast present). Y-axis indicates cell state, X-axis lists TFs. Differential expression (upregulated genes) is tested along invading trophoblast trajectory (as

shown in Fig. 2a) in a retrograde manner using limma approach (FDR < 0.05, with Bonferroni correction for multiple hypotheses testing. **b**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of signalling molecules upregulated in EVT (Y-axis) in all donors (n = 17 donors with trophoblast present). **c**: Heatmap showing z-score of normalised, log-transformed and scaled expression of TFs downregulated during trophoblast invasion in all donors (n = 17 donors with trophoblast present). Y-axis indicates cell state, X-axis lists TFs. Differential expression (downregulated genes) is tested along invading trophoblast trajectory (as shown in Fig. 2a) in a retrograde manner using limma approach (FDR < 0.05, with Bonferroni correction for multiple hypotheses testing). **d**: Schematic representation of signalling pathways in distinct spatial microenvironments. Villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), extravillous trophoblast (EVT), interstitial EVTs (iEVTs), giant cells (GC), endovascular EVT (eEVT), microenvironment (ME), transcription factors (TFs).

Extended Data Fig. 9 scRNA-seq and snRNA-seq data quality control and analysis overview of the trophoblast *in vitro* models.

a: Schematic representation of the extravillous trophoblast (EVT) differentiation experimental design, indicating time points and biological replicates in both primary trophoblast organoids (PTO, n = 6 scRNA-seq and n = 2 for snRNA-seq) and trophoblast stem cell (TSC, n = 2 for both scRNA-seq and snRNA-seq) models. **b**: Diagram showing the annotation of the *in vitro* models. Firstly, we analysed the datasets in four separate manifolds and annotated each of the cell states based on canonical markers. Secondly, we projected the trophoblast *in vivo* reference data onto the *in vitro* trophoblast subsets by building a logistic regression classifier that we trained on P13 snRNA-seq in vivo dataset. We excluded eEVT and GC cells as these are scarcely represented and our marker data show they are not present in the *in vitro* cultures. Thirdly, we integrated scRNA-seq data from *in vivo* and *in vitro* conditions using scVI, and used this manifold to calculate differentially expressed genes (DEG) amongst subsets. **c**: UMAP (uniform manifold approximation and projection) scatterplots of snRNA-seq (n = 3928 nuclei) of PTOs coloured by cell state, donor (n = 2 donors),

cell cycle phase and unbiased clustering using leiden. Sample integration was performed with Harmony. **d**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of main trophoblast subsets (Y-axis) in each of the clusters identified by unbiased clustering (a) in snRNA-seq of primary trophoblast organoids (PTOs). **e**: UMAP scatterplots of scRNA-seq of PTOs derived from n = 6 donors and coloured by donor, time-point, cell cycle phase, media cultured and unbiased clustering using leiden. **f**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of main trophoblast subsets in each of the clusters identified by unbiased clustering (c) in scRNA-seq of PTOs (Y-axis). **g**: Bar plots showing the proportion (%) of final cell states identified in data of each time point (left), media (center) and donor (right) for PTOs scRNA-seq. **h**: UMAP scatterplots of snRNA-seq (n = 1563 nuclei) of trophoblast organoids from trophoblast stem cells (TSC) coloured by cell state, donor, cell cycle phase and unbiased clustering using leiden. Sample integration was performed with Harmony. **i**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of main trophoblast subsets (Y-axis) in each of the clusters identified by unbiased clustering (f) in snRNA-seq of trophoblast stem cells (TSC). **j**: UMAP scatterplots of scRNA-seq of EVT derived from TSC coloured by donor, time-point, cell cycle phase and unbiased clustering using leiden. **k**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of main trophoblast subsets in each of the clusters identified by unbiased clustering (j) in scRNA-seq of trophoblast stem cells (TSC) (Y-axis). **l**: Bar plots showing the proportion (%) of final cell states identified in each donor's data (left) and time-point (right) for TSCs scRNA-seq. Trophoblast organoid media (TOM), syncytiotrophoblast (SCT), villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), proliferative (p), extravillous trophoblast (EVT), EVT media (EVTM), interstitial EVT (iEVT), giant cells (GC), endovascular EVT (eEVT), single-cell RNA sequencing (scRNA-seq), single-nuclei RNA sequencing (snRNA-seq).

Extended Data Fig. 10 Annotation and benchmark of trophoblast *in vitro* models.

a: Dot plots showing normalised, log-transformed and variance-scaled expression of genes (Y-axis) characteristic of trophoblast cell states (X-axis) in primary trophoblast organoids (PTO, left) and trophoblast stem cell (TSC, right) models. Marked in red are genes that differ from the *in vivo* observed expression pattern. **b**: Dot plot showing normalised, log-transformed and variance-scaled expression of genes (X-axis) of signalling molecules known to be upregulated in *in vivo* trophoblast invasion plotted in trophoblast organoids (Y-axis). **c**: UMAP (uniform manifold approximation and projection) scatterplot of PTO single-cell RNA sequencing (scRNA-seq data, n = 26,853 cells) coloured by predicted cell state (top left) and probability (top center) according to our logistic regression model. Only *in vivo* data from donor P13 (snRNA-seq) was considered for the training. Confusion matrix with predictions on test set based on common features with PTO (Top right). UMAP scatterplot of TSC scRNA-seq data (n = 9957 cells) coloured by predicted cell state (bottom left) and probability (bottom center) according to our logistic regression model. Only in vivo data from donor P13 (snRNA-seq) was considered for the training. Confusion matrix with predictions on test set based on common features with TSC (Bottom right). **d**: UMAP scatterplot of scRNA-seq data (n = 23,519 cells) from[13] re-annotated using markers from Fig. 1e. **e**: (Left and center) Integrated manifold (*in vivo* and *in vitro*) using scVI and coloured by cell state and specific conditions. Integration is performed with scVI. (Top right) Table displaying organoid-independent annotation for each scVI-integrated cluster. Organoid annotation matching *in vivo* labels displayed in green, discordant annotation in red. **f**: Violin plot showing normalised and log-transformed expression of differentially expressed genes (DEGs, limma, FDR < 0.05, with Bonferroni correction for multiple hypotheses testing) when comparing early EVT-2 in PTO vs *in vivo* EVT-2. (Right) Violin plot showing normalised and log-transformed expression of DEG when comparing iEVT in TSC vs *in vivo* iEVT. Primary trophoblast organoids (PTO), trophoblast stem cells (TSC), trophoblast organoid media (TOM), syncytiotrophoblast (SCT), villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), proliferative (p), extravillous trophoblast (EVT), interstitial EVT (iEVT), giant cells (GC), endovascular EVT (eEVT).

Extended Data Fig. 11 Predicted interactions between trophoblast and maternal immune cells.

a: UMAP (uniform manifold approximation and projection) scatterplot of single-cell RNA sequencing (scRNA-seq) and single-nuclei RNA sequencing (snRNA-seq) data of the 18 donors described in Extended Data Fig. 1c of the maternal-fetal interface (n = 325,665 cells and nuclei) coloured by cell state. Integration was performed with scVI. **b**: (Left) High-resolution imaging of a section of the placenta-decidua interface stained by smFISH for *HLA-G*, highlighting EVTs invading the decidua from the CCC. (Right) multiplexed co-staining with *NCAM1* (dNK marker), *CSF1* and cognate receptor *CSF1R*; dashed squares indicate areas shown magnified to right. (Bottom) solid and outlined arrows indicate neighbouring *CSF1R*-expressing EVTs and *CSF1*-expressing dNK cells, respectively. Representative image of samples from three donors. **c**: Dot plot showing normalised, log-transformed and variance-scaled gene expression of macrophage markers (X-axis) in data from (a) (Y-axis). **d**: High-resolution imaging of the placenta-decidua interface stained by multiplexed smFISH for *HLA-G* (EVTs), *EREG* (dM1), and *CD14* and *FOLR2* (dM2) for n = 4 donors (donor ID is specified in each panel). **e**: Dot plot showing normalised, log-transformed and variance-scaled expression of *CXCR6* (X-axis) on the EVT subsets present in TSC (n = 2). **f**: UMAP scatterplots of scRNA-seq of TSC (CXCL16 and BSA conditions) coloured by donor, cell cycle phase, time point, treatment and unbiased clustering using leiden (n = 2). **g**: Dot plot showing normalised, log-transformed and variance-scaled expression of marker genes of the main trophoblast subsets (X-axis) in cell clusters defined in (f) (Y-axis) from the integrated manifold of CXCL16 and BSA conditions in trophoblast stem cell (TSC) scRNA-seq (n = 2). **h**: UMAP scatterplot of scRNA-seq of TSC coloured by cell state (n = 2). Villous cytotrophoblast (VCT), cytotrophoblast cell column (CCC), proliferative (p), extravillous trophoblast (EVT), interstitial EVTs (iEVTs), giant cells (GC), endovascular EVT (eEVT), dendritic cells (DC), lymphatic (l), maternal (m), fetal (f), Hofbauer cells (HOFB), innate lymphocytes (ILC), macrophages (M), monocytes (MO), natural killer (NK), perivascular (PV), decidual (d), epithelial (epi), stromal (S), fibroblasts (F), uterine smooth muscle cells (uSMC), bovine serum albumin (BSA).

Extended Data Fig. 12 Interactions between trophoblast and perivascular (PV) cells.

a: Dot plot showing normalised, log-transformed and variance-scaled expression of perivascular (PV) cell state markers. **b**: UMAP (uniform manifold approximation and projection) scatterplot of scRNA-seq of PV cells (n = 2768 cells) coloured by the scaled gene expression of PV cell state markers. **c**: (Top) High-resolution imaging of adjacent sections of maternal-fetal interface stained by multiplexed smFISH for three gene panels, from two donors. Dashed squares indicate areas shown magnified underneath (middle and below), highlighting PV1-*AOC3*, PV1-*STEAP4*, and PV2-*MMP11* cells expressing each of their three respective marker genes. Solid arrows indicate relatively sparse PV1-*STEAP4* cells in second and fifth columns. **d**: (Top) High-resolution imaging of a section of decidua stained by smFISH for *HLA-G* (EVTs) multiplexed with *MYH11*, *FNDC1*, and *NTRK2* (PV1-AOC3); dashed squares indicate areas shown magnified below. (Middle) solid and outlined arrows indicate neighbouring PV1-AOC3 cells expressing *NTRK2* and EVTs, respectively. Representative image of samples from two donors. **e**: (Left) High-resolution imaging of a section of decidua stained by multiplexed smFISH for *HLA-G*, *NCAM1*, and *CXCL12*. Dashed squares highlight arteries containing *HLA-G*+ *NCAM1*+ eEVTs expressing *CXCL12*, shown magnified to right. Representative image of samples from two donors.

Supplementary information

Supplementary Information

This file contains a guide to Supplementary Tables 1–8 (tables supplied separately).

Reporting Summary

Supplementary Table 1

Sample metadata – see Supplementary Table Guide for details.

Supplementary Table 2

Quality control of samples for each 10X RNA library in our maternal-fetal interface atlas – see Supplementary Table Guide for details.

Supplementary Table 3

Annotation summary for each sample – see Supplementary Table Guide for details.

Supplementary Table 4

Variance explained (R2 column) in the MEFISTO model by each factor in each modality (RNA or ATAC) – see Supplementary Table Guide for details.

Supplementary Table 5

Differentially expressed genes (DEG) along trophoblast trajectories in P13 – see Supplementary Table Guide for details.

Supplementary Table 6

TF analysis along trophoblast trajectory – see Supplementary Table Guide for details.

Supplementary Table 7

Trophoblast interactions enriched by microenvironment (ME) using CellPhoneDB – see Supplementary Table Guide for details.

Supplementary Table 8

Probes used for multiplexed RNAscope smFISH.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Arutyunyan, A., Roberts, K., Troulé, K. *et al.* Spatial multiomics map of trophoblast development in early pregnancy. *Nature* **616**, 143–151 (2023). https://doi.org/10.1038/s41586-023-05869-0

- Received: 30 May 2022
- Accepted: 21 February 2023
- Published: 29 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05869-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

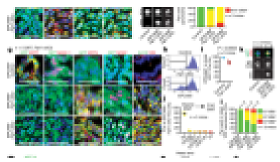
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

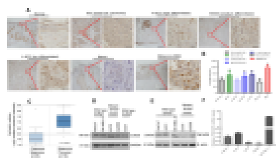
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05869-0



Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis

16 December 2020

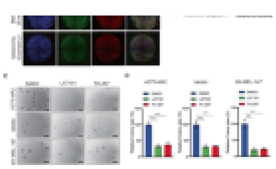
Ievgenia Pastushenko, Federico Mauri, … Cédric Blanpain



RETRACTED ARTICLE: MicroRNA-205-5p inhibits skin cancer cell proliferation and increase drug sensitivity by targeting TNFAIP8

11 March 2021

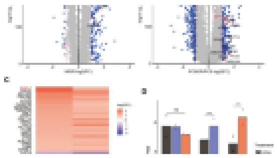
Xinhong Ge, Suryakant Niture, … Deepak Kumar



LIMK2 promotes melanoma tumor growth and metastasis through G3BP1-ESM1 pathway-mediated apoptosis inhibition

16 March 2023

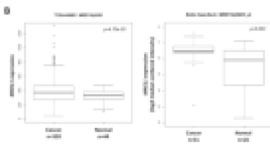
Parmanand Malvi, Dhana Sekhar Reddy, … Narendra Wajapeyee



[EGFL7 drives the evolution of resistance to EGFR inhibitors in lung cancer by activating NOTCH signaling]

29 October 2022

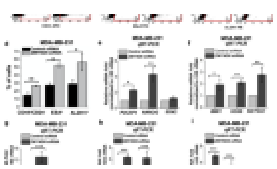
Yubo Wang, Pu Chen, … Mingwei Chen



[MRE11 promotes oral cancer progression through RUNX2/CXCR4/AKT/FOXA2 signaling in a nuclease-independent manner]

29 April 2021

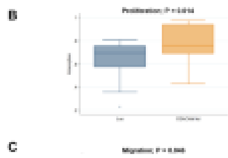
Yen-Yun Wang, Yuk-Kwan Chen, … Shyng-Shiou F. Yuan



[Suppression of poised oncogenes by ZMYND8 promotes chemo-sensitization]

15 December 2020

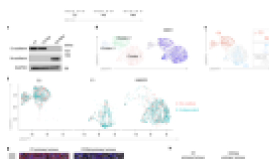
Shravanti Mukherjee, Santanu Adhikary, … Chandrima Das



[CEACAM1 promotes melanoma metastasis and is involved in the regulation of the EMT associated gene network in melanoma cells]

08 August 2018

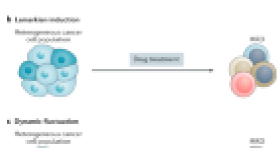
Daniel Wicklein, Benjamin Otto, … Udo Schumacher



[Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis]

11 April 2022

Yun Zhang, Joana Liu Donaher, … Robert A. Weinberg



[The great escape: tumour cell plasticity in resistance to targeted therapy]

10 October 2019

Soufiane Boumahdi & Frederic J. de Sauvage

- Article

- Open Access
- Published: 22 March 2023

RHOJ controls EMT-associated resistance to chemotherapy

- Maud Debaugnies,
- Sara Rodríguez-Acebes,
- Jeremy Blondeau,
- Marie-Astrid Parent,
- Manuel Zocco,
- Yura Song,
- Viviane de Maertelaer,
- Virginie Moers,
- Mathilde Latil,
- Christine Dubois,
- Katia Coulonval,
- Francis Impens,
- Delphi Van Haver,
- Sara Dufour,
- Akiyoshi Uemura,
- Panagiota A. Sotiropoulou,
- Juan Méndez &
- …
- Cédric Blanpain

Show authors

Nature **volume 616**, pages 168–175 (2023)

- 17k Accesses
- 1 Citations
- 161 Altmetric

- Metrics details

Abstract

The resistance of cancer cells to therapy is responsible for the death of most patients with cancer[1]. Epithelial-to-mesenchymal transition (EMT) has been associated with resistance to therapy in different cancer cells[2,3]. However, the mechanisms by which EMT mediates resistance to therapy remain poorly understood. Here, using a mouse model of skin squamous cell carcinoma undergoing spontaneous EMT during tumorigenesis, we found that EMT tumour cells are highly resistant to a wide range of anti-cancer therapies both in vivo and in vitro. Using gain and loss of function studies in vitro and in vivo, we found that RHOJ—a small GTPase that is preferentially expressed in EMT cancer cells—controls resistance to therapy. Using genome-wide transcriptomic and proteomic profiling, we found that RHOJ regulates EMT-associated resistance to chemotherapy by enhancing the response to replicative stress and activating the DNA-damage response, enabling tumour cells to rapidly repair DNA lesions induced by chemotherapy. RHOJ interacts with proteins that regulate nuclear actin, and inhibition of actin polymerization sensitizes EMT tumour cells to chemotherapy-induced cell death in a RHOJ-dependent manner. Together, our study uncovers the role and the mechanisms through which RHOJ acts as a key regulator of EMT-associated resistance to chemotherapy.

Main

Cancer is one of the main causes of mortality worldwide and resistance to therapy is responsible for treatment failure in the majority of patients[1]. EMT is a developmental process in which epithelial cancer cells lose cell–cell adhesion and acquire mesenchymal features, including increased invasiveness and motility[4]. EMT in cancer controls tumour initiation, progression, metastasis and resistance to anti-cancer therapies[2,5]. Transcription factors that promote EMT, such as Zeb1, Zeb2 and Twist1, mediate resistance to chemotherapy and radiotherapy in different cancer cell lines in vitro and after their transplantation into immunocompromised mice[6,7]. The conditional deletion of *Twist1* or *Snai1* in a mouse model of

pancreatic adenocarcinoma sensitizes tumour cells to gemcitabine[8]. EMT in a mammary tumour model promotes resistance to different chemotherapeutic drugs[9]. Several mechanisms have been proposed to account for the resistance to therapy associated with EMT in cancer cell lines in vitro[3,8,9]. However, whether these mechanisms are responsible for EMT-induced resistance to therapy in primary tumours in vivo remains unclear.

Here we used a genetic mouse model of squamous cell carcinoma (SCC) presenting spontaneous EMT to determine the functional response of different cell populations to chemotherapy. We found that, in primary skin SCCs presenting EMT, EPCAM^{-} tumour cells were highly resistant to chemotherapy. Using genetic gain and loss of function studies, we identified RHOJ, a small Rho GTPase, as a key regulator that promotes resistance to a broad range of chemotherapeutic agents in EMT tumour cells. Proteomic analysis combined with functional assays revealed that RHOJ controls the formation of nuclear F-actin fibres and the activation of dormant origins of replication, which promotes DNA repair and prevents chemotherapy induced cell death.

EMT-associated chemotherapy resistance in SCC

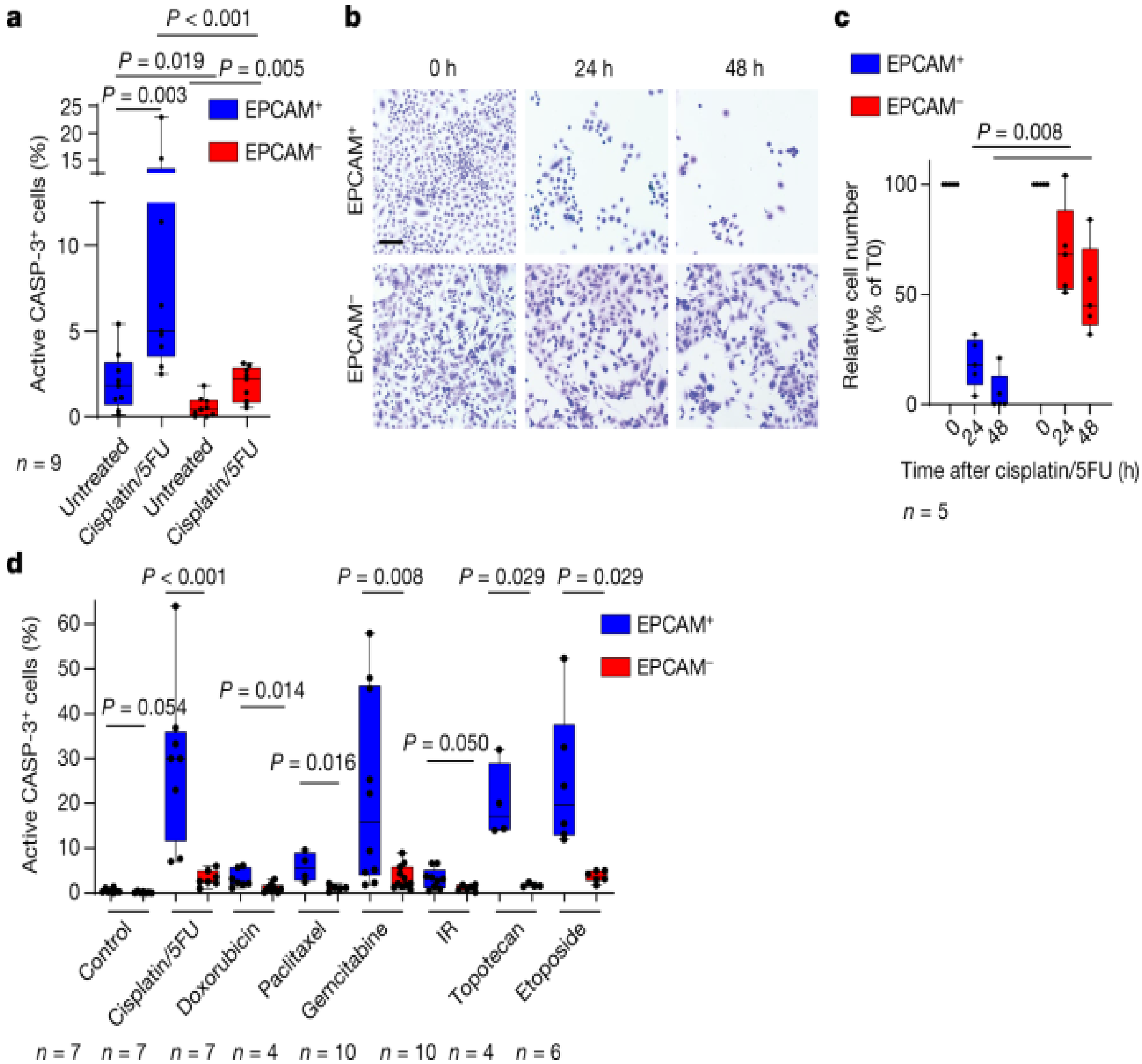
To investigate the role and the mechanisms of resistance to therapy in EMT tumour cells in vivo, we used a genetically induced mouse model of skin SCCs combining the expression of oncogenic *Kras* (*Kras*^{G12D}) with the deletion of *Trp53* and the expression of a YFP reporter in hair follicle lineages[10] (*Lgr5*^{creER}*Kras*^{G12D}*Trp53*^{cKO}*Rosa-YFP*). The YFP reporter, combined with EPCAM staining, enables us to identify tumour cells that undergo EMT and lose the expression of epithelial markers (YFP^{+}EPCAM^{-}) (Extended Data Fig. 1a).

To assess whether EMT is associated with resistance to chemotherapy in skin SCC, we treated *Lgr5*^{creER}*Kras*^{G12D}*Trp53*^{cKO}*Rosa-YFP* mice that were induced with tamoxifen and presented primary SCCs with cisplatin and 5-fluorouracil (5FU), the standard chemotherapy used to treat human patients with metastatic SCCs[11]. About 14% of SCCs were very sensitive to

chemotherapy and completely disappeared after 4 weeks of treatment (fully responding tumours). About 32% of SCCs did not respond and progressed during chemotherapy (non-responding tumours). Most SCCs (54%) presented a partial response to chemotherapy. Using immunostaining and fluorescence-activated cell sorting (FACS) analysis, we found that responding tumours consisted of YFP^+ epithelial tumour cells expressing the epithelial marker K14 or a mixture of epithelial and EMT tumour cells. The tumours that did not respond were enriched for EMT tumour cells ($K14^-VIM^+$ or $EPCAM^-$) (Extended Data Fig. 1b–j). These data suggest that chemotherapy either preferentially kills $EPCAM^+$ tumour cells or induces EMT promoting the switch from $EPCAM^+$ to $EPCAM^-$ tumour cells during chemotherapy.

To distinguish between these two possibilities, we investigated apoptosis in $EPCAM^+$ and $EPCAM^-$ tumour cells 24 h after chemotherapy administration. $EPCAM^-$ tumour cells were very resistant to chemotherapy-induced cell death in vivo as only 2% of $EPCAM^-$ tumour cells were caspase-3 positive compared with 9% of epithelial $EPCAM^+$ tumour cells. We recently identified EMT subpopulations of tumour cells presenting different degrees of EMT in mouse skin SCCs[12]. We found that the different EMT subpopulations were similarly resistant to chemotherapy (Fig. 1a and Extended Data Fig. 1k–p). Similar to what we found in vivo, EMT tumour cells were profoundly resistant to chemotherapy in vitro in the absence of the tumour microenvironment, whereas $EPCAM^+$ epithelial tumour cells were even more sensitive (Fig. 1b,c). In addition to cisplatin/5FU treatment, $EPCAM^-$ tumour cells were resistant to the topoisomerase I inhibitor topotecan, the topoisomerase II inhibitor etoposide and the anti-metabolite gemcitabine. $EPCAM^-$ tumour cells were also resistant to the anti-microtubule paclitaxel, the DNA intercalant and topoisomerase II inhibitor doxorubicin and ionizing radiation, although these treatments were less efficient in inducing cell death in $EPCAM^+$ tumour cells (Fig. 1d). Together, these data demonstrate that EMT in skin SCCs is associated with resistance to a wide range of genotoxic drugs in a cellular autonomous manner both in vivo and in vitro.

Fig. 1: EMT is associated with intrinsic resistance to chemotherapy in SCCs.



a, FACS quantification of the percentage of active caspase-3-positive cells in $YFP^{+}EPCAM^{+}$ and $YFP^{+}EPCAM^{-}$ cells from control and 24 h cisplatin/5FU-treated SCCs. *n* = 9 SCCs. **b**,**c**, Microscopy analysis of the cellular density of $EPCAM^{+}$ and $EPCAM^{-}$ tumour cells (**b**) and quantification of the number of cells (**c**) 24 h and 48 h after cisplatin/5FU administration in vitro. Scale bar, 100 µm. *n* = 5 independent experiments. **d**, FACS quantification of the percentage of activated caspase-3-positive cells in $YFP^{+}EPCAM^{+}$ and $YFP^{+}EPCAM^{-}$ cells 24 h after cisplatin/5FU administration (*n* = 7); doxorubicin (*n* = 7), paclitaxel (*n* = 4), gemcitabine (*n* = 10), topotecan (*n* = 4) or etoposide (*n* = 6) administration; and 24 h after receiving 10 Gy irradiation. *n* values represent the number of biological independent experiments. For **a**, **c** and **d**, statistical analysis was performed

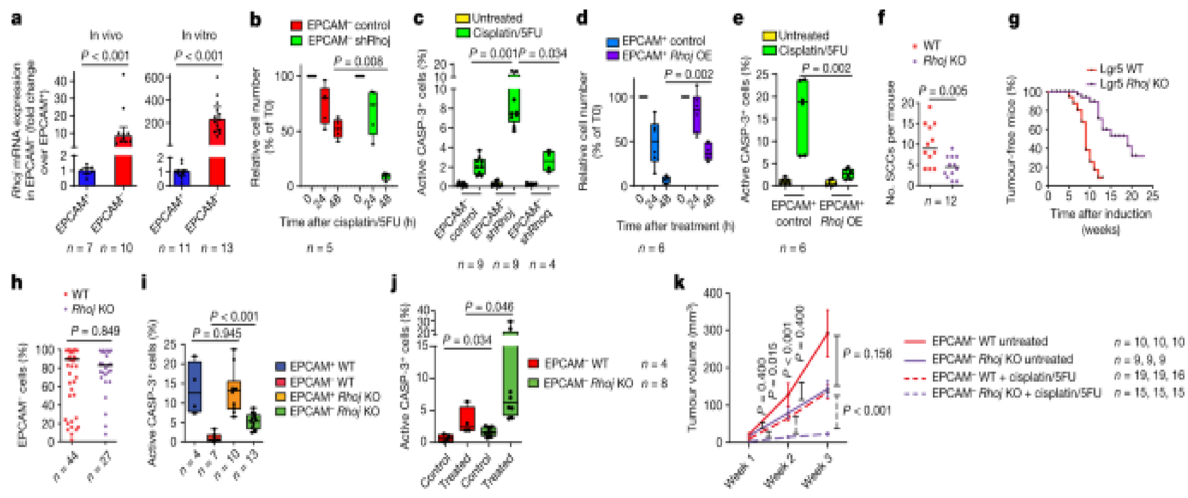
using two-sided Mann–Whitney *U*-tests. *P* values were adjusted for multiple comparisons using Bonferroni correction. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values.

Source data

RHOJ mediates EMT-associated resistance to therapy

To define the mechanisms by which EMT tumour cells are resistant to chemotherapy, we assessed the expression of putative mediators of resistance to therapy in $EPCAM^+$ and $EPCAM^-$ tumour cells from skin SCCs[10,12]. We found that EMT tumour cells expressed higher levels of apoptosis inhibitors (*Birc2*, *Birc3)*, glutathione-metabolism-related genes (*Gpx3*, *Gpx7*, *Gpx8*, *Gstm2*) and ABC multidrug transporters (*Abcb1a*, *Abcb1b*, *Abcg2*). We also found that RHOJ—a small GTPase of the Cdc42 subfamily that was previously described to mediate resistance to therapy in melanoma[13,14]—was expressed at a much higher level in EMT tumour cells (Fig. 2a and Extended Data Fig. 2a). To investigate whether RHOJ contributes to EMT-associated resistance to therapy in skin SCCs, we first assessed whether short hairpin RNA (shRNA)-mediated *Rhoj* knockdown (KD) sensitized EMT tumour cells to chemotherapy. KD of *Rhoj* but not *Rhoq*, its closest homologue, increased apoptosis after cisplatin/5FU treatment in EMT tumour cells, which was accompanied by a strong decrease in living tumour cells 48 h after chemotherapy. shRNA KD of *Rhoj* and *Rhoq* in $EPCAM^+$ tumour cells did not change their response to chemotherapy (Fig. 2b,c and Extended Data Fig. 2b–d). *Rhoj* KD in EMT tumour cells was associated with a slight decrease in cell growth in vitro and migration consistent with previous findings[15,16,17] and did not affect the expression of EMT or epithelial markers (Extended Data Fig. 2e–h).

Fig. 2: RHOJ mediates resistance to chemotherapy associated with EMT.



a, RT–qPCR analysis of *Rhoj* expression in EPCAM⁻ cells compared with in EPCAM⁺ cells in vivo (left) and in vitro (right). Data are median ± interquartile range, normalized to the housekeeping gene *Tbp*. *n* values indicate the number of independent experiments. **b**, The relative cell number of EPCAM⁻ control and *Rhoj* shRNA KD cells 24 h and 48 h after cisplatin/5FU administration in vitro. *n* values indicate the number of independent experiments. **c**, FACS quantification of the percentage of activated caspase-3-positive cells in EPCAM⁻ control, *Rhoj* shRNA KD and *Rhoq* shRNA KD cells 24 h after cisplatin/5FU administration. *n* values indicate the number of independent experiments. **d**,**e**, The relative cell number (**d**) and FACS quantification of the percentage of activated caspase-3-positive cells (**e**) of control and RHOJ-overexpressing (OE) EPCAM⁺ cells 24 h and 48 h after cisplatin/5FU administration in vitro. *n* = 6 independent experiments. **f**,**g**, The median number of SCCs per mouse (**f**) and kinetics of tumour appearance (**g**) in control and *Rhoj*-KO mice. *n* values indicate the number of mice. **h**, FACS quantification of the percentage of EPCAM⁻ cells of tumours from WT and *Rhoj*-KO SCCs. Data are median. *n* values indicate the number of mice. **i**,**j**, FACS quantification of the percentage of activated caspase-3-positive cells in EPCAM⁺ and EPCAM⁻ cells 24 h after cisplatin/5FU administration from WT (*n* = 4 and 7 SCCs) and *Rhoj*-KO mice (*n* = 10 and 13 SCCs) in vivo (**i**) and in EPCAM⁻ cells from WT-derived (*n* = 4) and *Rhoj*-KO-derived (*n* = 8) cell lines 24 h after cisplatin/5FU administration in vitro (**j**). *n* values indicate the number of independent experiments. **k**, The tumour volume after cisplatin/5FU administration to mice that were grafted with WT and *Rhoj*-KO tumour cells

compared with untreated mice. Data are mean ± s.e.m. n values indicate the number of mice. For **a**–**f** and **h**–**k**, statistical analysis was performed using two-sided Mann–Whitney U-tests. P values were adjusted using Bonferroni correction when multiple comparisons were performed. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values.

Source data

Next, we assessed whether RHOJ overexpression can confer resistance to chemotherapy in sensitive epithelial tumour cells. RHOJ overexpression in $EPCAM^+$ tumour cells strongly decreased the proportion of apoptotic tumour cells 24 h after cisplatin/5FU administration and leads to increased cell survival 48 h after chemotherapy (Fig. 2d,e and Extended Data Fig. 3a,b). The overexpression of RHOJ in $EPCAM^+$ tumour cells slightly decreased the growth of cells in vitro and did not affect the expression of EMT or epithelial markers (Extended Data Fig. 2h and 3c).

To determine whether RHOJ also mediates EMT resistance to therapy in vivo, we generated *Rhoj* conditional knockout (*Rhoj* KO) tumours using $Lgr5^{creER}Kras^{G12D}p53^{cKO}Rhoj^{cKO}Rosa\text{-}YFP$ mice (Extended Data Fig. 3d,e). *Rhoj* KO slightly decreased the number of tumours per mouse and increased their latency, suggesting that RHOJ promotes SCC initiation (Fig. 2f,g). The proportion of $EPCAM^+$ to $EPCAM^-$ tumour cells was unchanged after *Rhoj* deletion (Fig. 2h and Extended Data Fig. 3f), showing that RHOJ does not control EMT per se. *Rhoj*-KO tumour cells presented a decrease in proliferation as shown by Ki-67 immunostaining (Extended Data Fig. 3g,h). Administration of cisplatin/5FU to mice with tumours showed that *Rhoj*-KO tumours presented increased cell death in $EPCAM^-$ tumour cells (Fig. 2i and Extended Data Fig. 3i). Administration of chemotherapy in cultured cells in vitro showed that *Rhoj*-KO $EPCAM^-$ tumour cells were also more sensitive to cisplatin/5FU (Fig. 2j). To assess whether *Rhoj* deletion affects the long-term response of tumours to chemotherapy, we transplanted $EPCAM^-$ *Rhoj*-WT and *Rhoj*-KO tumour cells into immunodeficient mice and treated mice presenting growing tumours with chemotherapy. In the absence of chemotherapy, *Rhoj*-WT $EPCAM^-$ tumours grew a bit faster compared with *Rhoj*-KO $EPCAM^-$ tumours, consistent with their respective growth in vitro.

After cisplatin/5FU administration, *Rhoj*-WT EPCAM^{-} tumours continued to grow, although more slowly than in the absence of chemotherapy. By contrast, the size of *Rhoj*-KO EPCAM^{-} tumours was stable during the 3 weeks of cisplatin/5FU treatment, showing the short-term and long-term sensitization to chemotherapy after *Rhoj* deletion (Fig. 2k).

To assess the relevance of our findings to human cancers, we generated *RHOJ* KD using short hairpin RNA in MDA-MB-231 cells—a human triple-negative breast cancer cell line presenting EMT that expresses high levels of RHOJ and displaying p53 mutations. *RHOJ* KD in MDA-MB-231 cells decreased cell growth and increased cell death after administration of chemotherapy (Extended Data Fig. 4a–d).

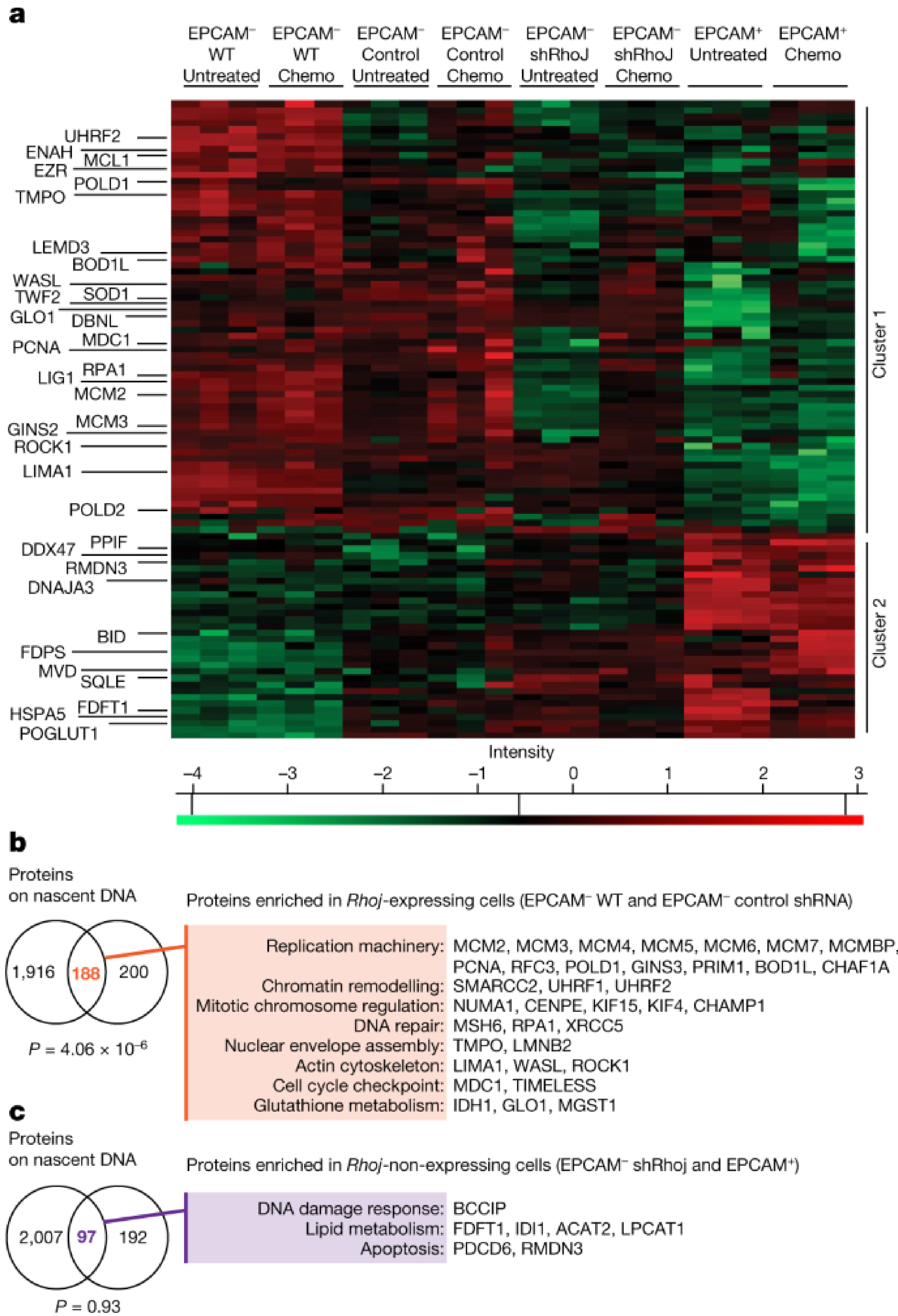
Transcriptome and proteome control by RHOJ

To define mechanistically how RHOJ promotes resistance to therapy in EMT tumour cells, we first performed bulk RNA-sequencing (RNA-seq) analysis of shRNA control and *Rhoj*-KD EPCAM^{-} tumour cells. When considering genes that were deregulated by twofold, 487 genes were downregulated, and 460 genes were upregulated by *Rhoj* KD in EPCAM^{-} cells. Gene Ontology analysis (GO) revealed that the genes downregulated in *Rhoj*-KD EPCAM^{-} cells are involved in DNA replication, cellular response to interferon and actin remodelling, whereas genes implicated in cell adhesion, lipid metabolism, apoptosis and microtubule and actin cytoskeleton were upregulated in *Rhoj*-KD EPCAM^{-} cells (Extended Data Fig. 5a–d).

In addition to the transcriptional regulation of the cellular effectors of DNA damage response (DDR) that accompanied genotoxic stress induced by chemotherapy, DDR factors are regulated at the protein level by protein stability, and post-translational modifications that control DNA repair and cell survival[18]. To define the effect of RHOJ on protein expression in EMT tumour cells, we performed proteomics analysis of EPCAM^{+} and EPCAM^{-} FACS-isolated primary tumour cells from *Lgr5^{creER}Kras^{G12D}p53^{cKO}Rosa-YFP* mice and YFP^{+}EPCAM^{-} cells that were transduced with viruses expressing control or *Rhoj* shRNA in untreated conditions and 24 h after

cisplatin/5FU administration. We identified the proteins of which the expression level was significantly regulated after *Rhoj* KD ($P < 0.05$, two-way analysis of variance (ANOVA)) and found that 681 proteins were significantly differentially regulated between RHOJ-expressing and non-expressing samples. *Rhoj*-KD tumour cells presented decreased expression of nuclear proteins, actin regulators, glutathione metabolism and proteins implicated in DNA replication. By contrast, proteins implicated in endoplasmic reticulum metabolic processes, cholesterol metabolism and apoptosis were upregulated in *Rhoj*-KD tumour cells (Fig. 3a and Extended Data Fig. 5e,f). Phalloidin staining showed an increase in the number of actin stress fibres after *Rhoj* KD, consistent with previous reports[16,19] (Extended Data Fig. 5g).Using western blotting, we validated the higher expression of proteins regulating DNA replication (POLD, PCNA) and repair (pRPA2) as well as actin cytoskeleton (N-WASP) in EPCAM^{-} tumour cells compared with in EPCAM^{+} tumour cells, and their downregulation in EPCAM^{-} *Rhoj*-KO cells (Extended Data Fig. 5h–k). The deletion of *Rhoj* in EPCAM^{-} tumour cells did not affect the expression of EMT or epithelial markers (Extended Data Fig. 5l,m).

Fig. 3: Proteomic analysis after *Rhoj* KO in EMT tumour cells.


a
EPCAM⁻ WT Untreated
EPCAM⁻ WT Chemo
EPCAM⁻ Control Untreated
EPCAM⁻ Control Chemo
EPCAM⁻ shRhoJ Untreated
EPCAM⁻ shRhoJ Chemo
EPCAM⁺ Untreated
EPCAM⁺ Chemo
UHRF2
ENAH
MCL1
EZR
POLD1
TMPO
LEMD3
BOD1L
WASL
TWF2
SOD1
GLO1
DBNL
PCNA
MDC1
LIG1
RPA1
MCM2
GINS2
MCM3
ROCK1
LIMA1
POLD2
DDX47
PPIF
RMDN3
DNAJA3
BID
FDPS
MVD
SQLE
HSPA5
FDFT1
POGLUT1
Cluster 1
Cluster 2
Intensity
–4
–3
–2
–1
0
1
2
3
b
Proteins on nascent DNA
1,916
188
200
P = 4.06 × 10⁻⁶
Proteins enriched in Rhoj-expressing cells (EPCAM⁻ WT and EPCAM⁻ control shRNA)
Replication machinery: MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MCMBP, PCNA, RFC3, POLD1, GINS3, PRIM1, BOD1L, CHAF1A
Chromatin remodelling: SMARCC2, UHRF1, UHRF2
Mitotic chromosome regulation: NUMA1, CENPE, KIF15, KIF4, CHAMP1
DNA repair: MSH6, RPA1, XRCC5
Nuclear envelope assembly: TMPO, LMNB2
Actin cytoskeleton: LIMA1, WASL, ROCK1
Cell cycle checkpoint: MDC1, TIMELESS
Glutathione metabolism: IDH1, GLO1, MGST1
c
Proteins on nascent DNA
2,007
97
192
P = 0.93
Proteins enriched in Rhoj-non-expressing cells (EPCAM⁻ shRhoj and EPCAM⁺)
DNA damage response: BCCIP
Lipid metabolism: FDFT1, IDI1, ACAT2, LPCAT1
Apoptosis: PDCD6, RMDN3

a, The relative intensities of selected upregulated proteins identified in RHOJ-expressing cells (EPCAM^{-} WT and EPCAM^{-} control shRNA) and in RHOJ-non-expressing cells (EPCAM^{+} and EPCAM^{-} *Rhoj* shRNA KD) in untreated conditions and after 24 h of cisplatin/5FU treatment. Two major clusters were observed. $n = 3$ technical replicates. *Z*-scored $\log_2$-transformed label-free quantitation (LFQ) intensities are shown. Statistical analysis was performed using two-way ANOVA; $P < 0.05$. Individual *P* values are provided as Source Data. **b**,**c**, The overlap between a published proteomic study of proteins associated with replicative fork (iPond)[20] and the proteins enriched in RHOJ-expressing cells (**b**) or the proteins enriched in RHOJ-non-expressing cells (**c**) in the untreated condition and 24 h after cisplatin/5FU administration. The *P* value was calculated using the hypergeometric test.

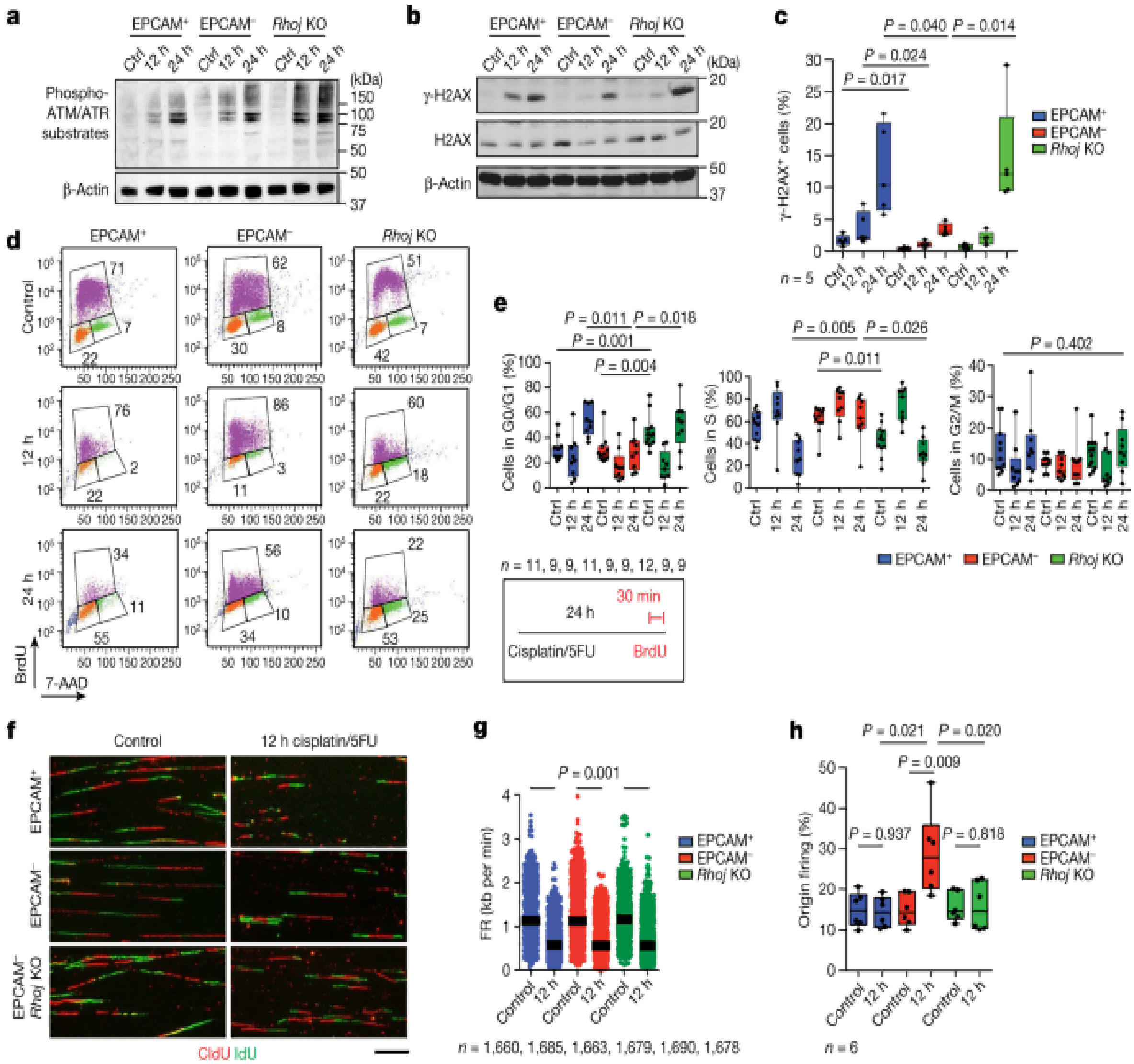
Source data

Importantly, the proteins differentially regulated by *Rhoj* KD were strongly enriched for proteins found on newly replicated DNA after replicative stress (iPOND)[20] (Fig. 3b,c). These included several replisome components, such as the helicase MCM2–7; different subunits of the DNA polymerases and their accessory factors (such as PCNA, PRIM1, POLD1); components of the PCNA-loading replication factor C (RFC3); protective factors of stalled replication forks (such as BODL1 and TIMELESS); initiation factors (GINS3); and chromatin assembly factors (CHAF1A). Other proteins identified in the proteomics assays included chromatin-remodelling factors (SMARCC2, UHRF1, UHRF2); intra-S-phase checkpoint signalling mediator (MDC1); spindle-associated proteins (such as CHAMP1, KIF15, KIF4, NUMA1, CENPE); proteins implicated in nuclear envelope assembly, including LEM protein (TMPO) and lamin (LMNB2); established RHOJ partners (WASL and ROCK1)[21]; single-strand DNA-binding protein RPA1; and different DNA repair factors, such as the core-component of non-homologous end-joining XRCC5 and the component of post-replicative DNA mismatch repair MSH6. The important overlap between proteins regulated by RHOJ expression and proteins associated with nascent DNA and/or recruited during replicative stress suggests that RHOJ controls DNA replication and DNA repair.

RHOJ promotes DNA repair in EMT tumour cells

To assess whether chemotherapy in $EPCAM^+$ and $EPCAM^−$ as well as $EPCAM^−$ *Rhoj*-KO tumour cells induces DNA damage and differentially activates the DDR pathways, we investigated the level of phosphorylation of ATM and ATR substrates, two main kinases that activate the DDR after DNA damage[22]. $EPCAM^+$, $EPCAM^−$ and $EPCAM^−$ *Rhoj*-KO tumour cells similarly increased the level of phosphorylated ATM/ATR substrates after cisplatin/5FU administration (Fig. 4a), showing that RHOJ does not control ATM and ATR activation after chemotherapy. ATR and ATM inhibitors did not sensitize $EPCAM^−$ cells to chemotherapy (Extended Data Fig. 6a–c), suggesting that $EPCAM^−$ resistance to therapy is independent of ATR- and ATM-mediated checkpoint. To determine whether RHOJ inhibits EMT tumour cells apoptosis by promoting DNA repair and/or replicative tolerance to damaged DNA, as suggested by the proteomic data, we first investigated the level of histone 2A.X phosphorylated at Ser139 (γ-H2AX), a mark of DNA damage after chemotherapy. Western blot, immunofluorescence and FACS quantification showed that $EPCAM^−$ cells presented a significant reduction in γ-H2AX compared with $EPCAM^+$ and $EPCAM^−$ *Rhoj*-KO tumour cells after cisplatin/5FU administration (Fig. 4b,c and Extended Data Fig. 6d,e), supporting the notion that RHOJ prevents accumulation of DNA damage in EMT tumour cells.

Fig. 4: RHOJ promotes DNA repair and the activation of new origins of DNA replication after chemotherapy in EMT tumour cells.



a, Western blot analysis of phosphorylation of ATM/ATR substrates in EPCAM^{+}, EPCAM$^{−}$ and *Rhoj*-KO EPCAM$^{−}$ cells untreated and treated with cisplatin/5FU for 12 h and 24 h. $n = 3$. **b**,**c**, Western blot analysis (**b**; $n = 3$) and FACS analysis quantification (**c**; $n = 5$) of the percentage of cells expressing γ-H2AX in EPCAM^{+}, EPCAM$^{−}$ and EPCAM$^{−}$ *Rhoj*-KO cells, 12 h and 24 h after cisplatin/5FU administration. β-Actin was the loading control. *n* values represent the number of independent primary cultured cell lines. **d**,**e**, Representative cell cycle FACS profile (**d**) and quantification (**e**) of the percentage of cells in G0/G1, S and G2/M phase 12 h and 24 h after cisplatin/5FU treatment of EPCAM^{+}, EPCAM$^{−}$ and EPCAM$^{−}$ *Rhoj*-KO BrdU-7AAD-labelled cells. *n* values indicate the number of independent experiments. **f**, Representative stretched DNA fibres labelled with CldU and

IdU from EPCAM$^+$, EPCAM$^-$ and EPCAM$^-$ *Rhoj*-KO cells untreated and treated with cisplatin/5FU for 12 h. $n = 6$. Scale bar, 10 µm. **g**, Fork rate (FR) values in EPCAM$^+$, EPCAM$^-$ and EPCAM$^-$ *Rhoj*-KO cells untreated and treated with cisplatin/5FU for 12 h. Data are median. Statistical analysis was performed using ANOVA, with condition, experience and their interaction; the *P* value was adjusted using the two-way post hoc Sidak test. *n* values represent the number of forks pooled from six independent experiments. **h**, The percentage of origin of firing in EPCAM$^+$, EPCAM$^-$ and EPCAM$^-$ *Rhoj*-KO cells untreated and treated with cisplatin/5FU for 12 h. From left to right, total number of DNA fibres scored: $n = 3{,}020$, 3,123, 3,065, 3,091, 3,078 and 3,069, in six independent experiments. For **c**, **e** and **h**, statistical analysis was performed using Kruskal–Wallis tests followed by two-sided Mann–Whitney *U*-tests; *P* values were adjusted for multiple comparisons using Bonferroni correction. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values.

To assess at which step of the DDR pathway RHOJ reduces the extent of DNA damage or regulates DNA repair, we analysed the expression and subcellular localization of key mediators of the DDR (53BP1, RPA2 and RAD51) after cisplatin administration using immunofluorescence microscopy. Clustering of 53BP1 at the site of DNA damage is an early event in the DDR[23] and acts together with other cofactors including RIF1 to block the over-resection of DNA at the break[24]. We found that, 12 h after cisplatin administration, 53BP1 nuclear foci were more abundant in EPCAM$^-$ tumour cells compared with in EPCAM$^+$ and EPCAM$^-$ *Rhoj*-KD cells. RPA2 focus formation is associated with single-stranded DNA generated by DNA end-resection at damage sites and stalled replication forks after replicative stress[25,26]. RAD51 focus formation after DNA damage is associated with homologous-recombination-mediated DNA repair of double-stranded breaks and is recruited at stalled forks generated during replicative stress to mediate fork reversal and protect the DNA from nucleolytic degradation[27]. The number of RPA2 and RAD51 nuclear foci was lower in EPCAM$^-$ cells compared with in EPCAM$^+$ and EPCAM$^-$ *Rhoj*-KD cells (Extended Data Fig. 6f). These results are consistent with an increase in DNA repair and a decrease in replicative stress in the presence of RHOJ in EPCAM$^-$ tumour cells, in good accordance with the γ-H2AX data.

RHOJ actives new origins of DNA replication

To determine whether the cell cycle is differentially regulated in EPCAM^{+}, EPCAM$^{−}$ and EPCAM$^{−}$ *Rhoj*-KD tumour cells in response to chemotherapy, we assessed DNA content and BrdU incorporation using BrdU/7AAD FACS analysis. The cell cycle profile of EPCAM^{+} and EPCAM$^{−}$ cells was comparable in the untreated condition, whereas *Rhoj*-KO EPCAM$^{−}$ tumour cells presented an increase in G0/G1 at the expense of S phase. At 12 h after cisplatin/5FU treatment, the proportion of BrdU-positive cells increased in all of the conditions. However, 24 h after cisplatin/5FU administration, the percentage of cells in S phase was strongly reduced in EPCAM^{+} and *Rhoj*-KO tumour cells, whereas EPCAM$^{−}$ tumour cells continued to synthesize DNA, as revealed by BrdU incorporation (Fig. 4d,e). Genotoxic agents cause DNA damage in replicative cells by increasing replicative stress, which in turn activates intra-S-phase checkpoint pathways to facilitate replication completion and avoid double-stranded break formation arising from fork collapse, thereby limiting cell death[28]. The level of EdU incorporation was greatly reduced in all tumour cells 24 h after cisplatin/5FU administration, consistent with the activation of the intra-S-phase checkpoint (Extended Data Fig. 7a,b). ATR kinase is a major regulator of the replicative stress checkpoint, whereby it targets different proteins—including CHK1 and WEE1, which control cyclin-dependent kinase (CDK) activity—to activate cell cycle arrest in S phase and prevent mitotic catastrophe[29]. Higher activation of CDKs, including CDK1 and CDK4, was observed in EPCAM$^{−}$ tumour cells compared with in EPCAM^{+} and *Rhoj*-KO EPCAM$^{−}$ cells after chemotherapy (Extended Data Fig. 7c–f) suggesting that RHOJ enables EMT tumour cells to progress in the cell cycle and continue DNA synthesis after chemotherapy.

To test whether the higher resistance of EPCAM$^{−}$ tumour cells to chemotherapy is related to the higher ability to cope with replicative stress, we evaluated the sensitivity of EPCAM^{+}, EPCAM$^{−}$ and *Rhoj*-KO tumour cells to the replicative-stress-inducing drug aphidicolin. Aphidicolin is a selective inhibitor of DNA polymerases that induces fork stalling and single-strand DNA exposure by uncoupling polymerase and helicase activities[30]. Aphidicolin induced a major increase in EPCAM^{+} and *Rhoj*-KO EPCAM$^{−}$

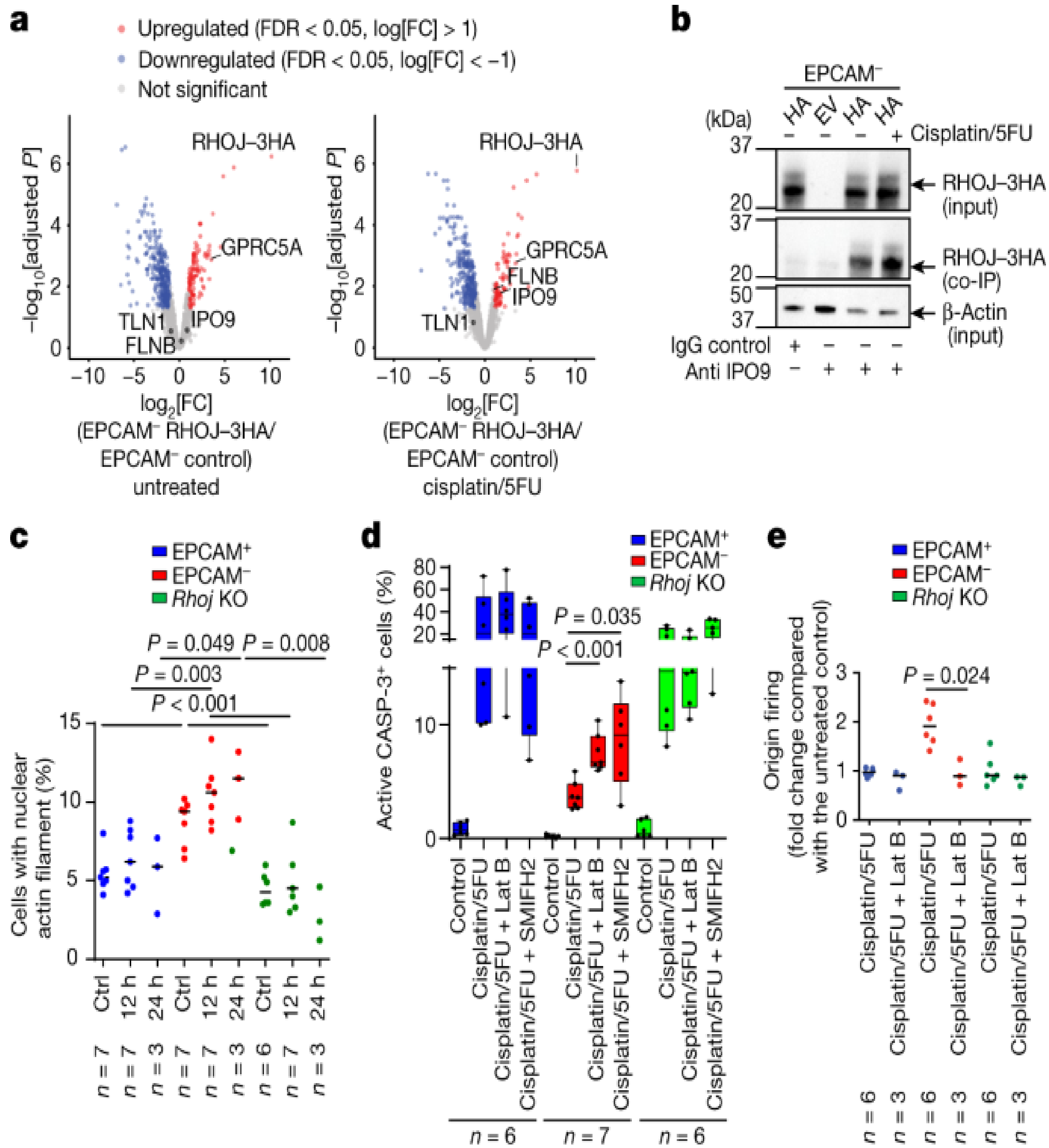
cell death. By contrast, EPCAM^{-} tumour cells were relatively insensitive to aphidicolin treatment (Extended Data Fig. 7g). In the case of insufficient protection of stalled forks during replicative stress, the nascent DNA will be degraded by different nucleases (for example, DNA2 and MRE11), leading to fork collapse[31]. Treatment with mirin, a MRE11 inhibitor, after cisplatin/5FU administration decreased cell death in EPCAM^{+} and *Rhoj*-KO EPCAM^{-} cells but had no effect on EPCAM^{-} tumour cells (Extended Data Fig. 7h), suggesting that the absence of RHOJ leads to replicative stress after chemotherapy. There was an increase in the number of micronuclei in EPCAM^{+} and *Rhoj*-KO EPCAM^{-} cells compared with in EPCAM^{-} cells 12 h and 24 h after chemotherapy (Extended Data Fig. 8a,b), indicating unresolved replicative stress in the absence of RHOJ.

To understand how RHOJ promotes continued DNA replication (Fig. 4d,e) and DDR, we assessed the speed of replication forks and the activation of replication origins under basal conditions and after chemotherapy[32]. Using DNA fibre analysis after sequential pulse of CldU and IdU, we found that cisplatin/5FU treatment induced a global slowing of replication forks in all cell types, further demonstrating the activation of the S-phase checkpoint in all conditions (Fig. 4f,g). The activation of the S-phase checkpoint was the consequence of fork stalling, as shown by the increase in the asymmetry of fork progression in all cell types after chemotherapy (Extended Data Fig. 8c,d). By contrast, the number of origins was increased only in EPCAM^{-} cells as compared with EPCAM^{+} and *Rhoj*-KO EPCAM^{-} cells in response to cisplatin/5FU treatment, showing that chemotherapy induced the firing of new origins of replication in EPCAM^{-} cells (Fig. 4h). Dormant DNA replication origins are activated to resume replication in case of fork stalling during replicative stress. It has been demonstrated that the amount of MCM proteins affects the availability of dormant origins[33,34]. Our data show that EPCAM^{-} cells display a higher concentration of several MCM subunits on chromatin (Extended Data Fig. 8e), consistent with the ability of EPCAM^{-} tumour cells to promote the firing of dormant origins in response to chemotherapy.

RHOJ regulates nuclear actin polymerization

To understand how RHOJ regulates DNA repair and the activation of origins of DNA replication after chemotherapy, we sought to identify the proteins that interact with RHOJ in EMT tumour cells. To this end, we performed affinity purification followed by mass spectrometry (AP–MS) using anti-HA pull-down of $EPCAM^{-}$ cells expressing 3HA-tagged RHOJ under untreated conditions and 12 h after cisplatin/5FU administration. We identified 118 proteins that were significantly enriched in untreated HA-tagged-RHOJ-expressing cells and 98 proteins that were significantly enriched in treated HA-tagged-RHOJ-expressing cells. Among them, we found previously identified RHOJ-interacting proteins such as GPRC5A[21]. Notably, after chemotherapy treatment, there was an enrichment of proteins implicated in the regulation of actin filament dynamics (FLNB and TLN1) and IPO9—a protein that regulates nuclear import of actin[35] (Fig. 5a). Immunostaining of HA-tagged-RHOJ-expressing tumour cells showed the cytoplasmic, perinuclear and nuclear localization of RHOJ (Extended Data Fig. 9a). These findings suggest that RHOJ physically interacts with nuclear actin regulators. Nuclear actin has been shown to regulate DNA replication and repair[36,37,38,39,40,41] and KD of *IPO9* has been shown to decrease nuclear actin filaments and impair DNA repair[37]. Co-immunoprecipitation experiments confirmed the presence of IPO9 in the RHOJ protein complex (Fig. 5b), suggesting that RHOJ regulates the response to chemotherapy through the modulation of nuclear actin function.

Fig. 5: RHOJ promotes cell survival and DNA repair by regulating nuclear actin polymerization.



a, Differential protein expression detected by AP–MS analysis of protein complexed with 3HA-tagged RHOJ in EPCAM^{-} cells. Volcano plot showing the statistical significance (−$\log_{10}$-transformed *P*) and the fold change ($\log_2$) of proteins identified by MS (false-discovery rate (FDR) < 0.05) in the untreated condition (left) and after 12 h of cisplatin/5FU treatment (right). Selected proteins enriched after immunoprecipitation of 3HA-tagged RHOJ are highlighted. **b**, Co-immunoprecipitation using anti-IPO9 or IgG control antibodies followed by western blotting using anti-HA antibodies, revealing

the association between IPO9 and RHOJ–3HA protein in EPCAM⁻ cells 12 h after cisplatin/5FU treatment. $n = 2$. **c**, Quantification of the percentage of cells presenting nuclear actin filaments in $EPCAM^+$, $EPCAM^-$ and $EPCAM^-$ *Rhoj*-KO cells expressing nuclear actin chromobody-GFP 12 h and 24 h after cisplatin/5FU treatment (>250 nuclei were quantified per replicate). Data are median. Statistical analysis was performed using ANOVA, with condition, experience and their interaction. *P* values were adjusted using two-way post hoc Sidak tests. *n* values indicate the number of replicates. **d**, FACS quantification of the percentage of activated caspase-3-positive cells in $EPCAM^+$, $EPCAM^-$ and $EPCAM^-$ *Rhoj*-KO cells 24 h after treatment with cisplatin/5FU and latrunculin B (Lat B) or SMIFH2 inhibitor. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values. *n* values indicate the number of replicates. **e**, The relative proportion of origin firing in $EPCAM^+$, $EPCAM^-$ and $EPCAM^-$ *Rhoj*-KO cells treated with cisplatin/5FU or a combination of cisplatin/5FU and latrunculin B for 12 h compared with the untreated condition. Data are median. *n* values indicate the number of independent experiments. For **d** and **e**, statistical analysis was performed using Kruskal–Wallis tests followed by two-sided Mann–Whitney *U*-tests. *P* values were adjusted for multiple comparisons using Bonferroni correction. Gel source data are provided in Supplementary Fig. 1.

Source data

To test this possibility, we investigated the presence of actin fibres in the nucleus using transfection of nuclear actin chromobody in $EPCAM^+$, $EPCAM^-$ and *Rhoj*-KO tumour cells isolated from $Lgr5^{creER}Kras^{G12D}Trp53^{cKO}$*Rosa-tdTomato* mice. The proportion of cells with nuclear actin fibres was increased in $EPCAM^-$ tumour cells compared with in $EPCAM^+$ and *Rhoj*-KO tumour cells in the presence or absence of cisplatin/5FU, showing that RHOJ controls nuclear actin polymerization (Fig. 5c). Different patterns of nuclear actin filaments were identified. After chemotherapy, there was an increase in elongated actin filaments in $EPCAM^-$ cells compared with the short and branched actin filaments observed in $EPCAM^+$ cells. *Rhoj*-KO $EPCAM^-$ cells were associated with short filaments with a dense and multipolar organization (Extended Data

Fig. 9b,c). Immunostaining of nuclear actin with phosphorylated histone H3 and EdU showed that chemotherapy promoted the expression of nuclear actin in replicative cells in a RHOJ-dependent manner (Extended Data Fig. 10a–d).

To investigate whether nuclear actin polymerization promotes RHOJ-mediated replication stress tolerance, we assessed the effects of the inhibition of nuclear actin polymerization on the replicative-stress response. A combination of cisplatin/5FU treatment with the F-actin inhibitor latrunculin B significantly and specifically increased apoptosis in EPCAM^{-} cells (Fig. 5d and Extended Data Fig. 10e). Latrunculin B treatment inhibited the formation of new origins of replication and the disappearance of RPA2 foci after chemotherapy in EPCAM^{-} cells (Fig. 5e and Extended Data Fig. 10f), suggesting that RHOJ-mediated replication-stress tolerance is dependent on nuclear actin polymerization. Among different inhibitors of actin remodelling, only the formin inhibitor SMIFH2 increased apoptosis in EPCAM^{-} cells when combined with chemotherapy treatment (Fig. 5d and Extended Data Fig. 10g). Together, these data indicate that RHOJ controls DNA repair by regulating nuclear actin polymerization, which, in turn can modulate the processes of DNA replication and repair.

Discussion

Here we identified a molecular mechanism by which EMT tumour cells resist a broad range of genotoxic and chemotherapeutic agents in a primary mouse model of skin SCC presenting spontaneous EMT. We found that RHOJ, a small Rho GTPase, is overexpressed in tumour cells presenting EMT compared with in epithelial tumour cells. We found, through genetic gain and loss of function studies in vitro and in vivo, the key role of RHOJ in regulating the resistance of EMT tumour cells to chemotherapy. Whereas small GTPases of the Rho family have been previously associated with DNA repair and response to therapy[42,43], the role of RHOJ in regulating resistance to therapy in EMT tumour cells was unclear.

RHOJ modulated the response to chemotherapy in melanoma cell lines by suppressing ATR and CHK1 activation and decreasing p53-mediated cell death[13]. By contrast, p53 is deleted in skin SCC presenting EMT, and

inhibition of ATR does not affect the resistance of EMT tumour cells to chemotherapy, suggesting that RHOJ regulates other mechanisms to mediate resistance to therapy in EMT tumour cells.

RHOJ controls several aspects of tumorigenesis, including tumour initiation and tumour growth, by regulating the rate of tumour cell proliferation. Moreover, RHOJ inhibits cell death after the administration of distinct types of chemotherapy and DNA-damaging agents, and the cells are repaired by different pathways, suggesting that RHOJ does not regulate a specific DNA-repair pathway but, rather, controls a common mechanism shared by the different DDR pathways and mechanisms counteracting replication stress. Our proteomic analysis shows that RHOJ regulates the expression of many proteins that promote DNA replication and DNA repair. After DNA damage and replication stress induced by chemotherapy, both epithelial and EMT tumour cells decrease the speed of DNA replication by activating the intra-S checkpoint. However, only EMT tumour cells activate dormant origins in a RHOJ-dependent manner, which facilitates continued DNA replication despite the slowdown of replication forks and promotes cell survival after chemotherapy.

EMT is associated with remodelling of the actin cytoskeleton, which is essential to promote cell migration and invasion[44]. Although RHOJ was not previously associated with changes in actin remodelling occurring during EMT, it has been implicated in the regulation of actin polymerization, the formation of localized actin stress fibres, and the reduction in actomyosin contractility in fibroblasts and endothelial cells[16,19]. RHOJ also promotes TGF-β-induced breast cancer cell migration and invasion[15] and regulates melanoma cell migration through the regulation of the actin cytoskeleton[45]. RHOJ controls neurite extension by binding to and activating N-WASP to induce ARP2/3-complex-mediated actin polymerization[46]. RHOJ also interacts in a complex containing other actin regulators, including formin like 1 and WASPL/N-WASP[21].

Notably, recent studies have implicated nuclear actin in DNA repair and replication fork repair. Inactivation of ARP2/3, formin or nuclear actin polymerization decrease DNA repair[36,37,39,41,47] and formin has been shown to promote the initiation of DNA replication through the regulation of

nuclear actin dynamics[40]. Our data demonstrate that RHOJ promotes resistance to therapy in EMT tumour cells by modulating DNA repair and activating dormant replication origins through the regulation of formin-dependent nuclear actin polymerization.

In conclusion, our study identified the key role of RHOJ and the mechanisms by which RHOJ promotes resistance to chemotherapy in EMT tumour cells, with important implications for the development of new strategies to overcome resistance to anticancer therapy.

Methods

Mice

Rosa26-YFP[48], *Rosa-tDTomato*[49], *K14*^{creER}, *Lgr5*^{creER} (ref. [50]), *KrasLSL*^{G12D} (ref. [51]) and *Trp53*$^{fl/fl}$ (ref. [52]) mice were obtained from the NCI mouse repository and Jackson Laboratories. *Rhoj*$^{fl/fl}$ mice[53] were a gift from A. Uemura. All mice used in this study were composed of males and females with mixed genetic background. Mouse colonies were maintained in a certified animal facility in accordance with the European guidelines. The room temperature ranged from 20 °C to 25 °C. The relative ambient humidity at the level of mouse cages was 55 ± 15%. Each cage was provided with food, water and two types of nesting material. A semi-natural light cycle of 12 h–12 h light–dark was used. All of the experiments were approved by the Ethical Committee for Animal Welfare (Commission d' Ethique et du Bien Etre Animal, CEBEA, Faculty of Medicine, Université Libre de Bruxelles, reference no. 434N and 663N). Sample sizes for experiments involving mice were determined according to protocols 434N and 663N, stating the number of mice used for experiments should be reduced to the minimum as soon as the result is reproducible within each experiment.

Kras ^{G12D} *Trp53* $^{fl/fl}$-induced skin tumours

Tamoxifen was diluted at 25 mg ml^{-1} in sunflower oil (Sigma-Aldrich). Tamoxifen (2.5 mg) was administered intraperitoneally for 4 days to

*Lgr5^creER^KrasLSL^G12D^Trp53^fl/fl^Rosa-YFP^+/+^* mice at postnatal day 28 as previously described[52,54].

Monitoring of tumour growth

Tumour appearance and size were detected by daily observation and palpation. Mice were euthanized when the tumour size reached more than 1 cm of diameter or when the mice presented signs of distress or lost more than 15% of their initial weight as permitted by our IACUC. These limits were not exceeded in any of the experiments. Skin tumours were measured using precision callipers enabling us to discriminate size modifications of greater than 0.1 mm. Tumour volumes were measured on the first day of appearance of the tumour and then every week until the death of the animal using the formula $V = D \times d \times h \times \pi/6$, where d is the minor tumour axis, D is the major tumour axis and h is the height. No macroscopic or microscopic intestinal tumours were observed in WT ($Lgr5^{creER}Kras^{G12D}p53^{cKO}Rosa\text{-}YFP$) or *Rhoj*-KO ($Lgr5^{creER}Kras^{G12D}p53^{cKO}Rhoj^{cKO}Rosa\text{-}YFP$) mice harbouring skin SCC that could have influenced general health and chemotherapy response.

Primary cell culture

FACS-isolated tumour YFP^+EPCAM^+ or YFP^+EPCAM^- cells were plated on γ-irradiated 3T3 feeder cells in six-well plates. Cells were cultured in MEM medium supplemented with 10% fetal bovine serum (FBS), 0.4 mg ml^{-1} hydrocortisone, 2×10^{-9} M T3, 1% penicillin–streptomycin and 2 mM l-glutamine. The feeders were removed using PBS/EDTA (1 mM). Cells were incubated at 37 °C with 20% O_2 and 5% CO_2.

Human cell culture

The MDA-MB-231 cell line (ATCC HTB-26) was grown in Dulbecco's modified Eagle's medium supplemented with 10% FBS and 1% penicillin–streptomycin. Cells were incubated at 37 °C with 20% O_2 and 5% CO_2. No commonly misidentified cell lines were used in this study according to

ICLAC register version II. All cell lines have been tested negative for mycoplasma contamination.

Chemotherapy treatments, irradiation and inhibitors

Mice were treated with 3.5 mg per kg *cis*-diammineplatinum dichloride (cisplatin) (Sigma-Aldrich, P4394) and 15 mg per kg 5FU (Sigma-Aldrich, F6627) administered intraperitoneally weekly for 4 weeks. For in vivo studies of primary mouse models, animals were selected according to their correct genotypes. The mice were induced with tamoxifen injection 28–35 days after birth and developed tumours in 2–3 months, thus minimizing the difference in age of different animals used. When tumour sizes reached 2–5 mm^3, mice were treated with chemotherapy injected intraperitoneally; their tumours were compared with those developing in mice of the same genotype injected with physiological serum.

Mouse tumoural cells were plated on six-well plates. For the irradiation, 10 Gy of radiation was delivered from a ^{137}Cs source, at a dose rate of 2.34 Gy min^{-1}. Cells were collected 24 h after the administration of the dose. For the chemotherapy, cells were incubated with 8.5 µM cisplatin (Sigma-Aldrich, P4394), 15 µM 5FU (Sigma-Aldrich, F6627), 0.4 µM gemcitabine (Sigma-Aldrich, G6423), 200 nM paclitaxel (Sigma-Aldrich, T7402) and 0.25 µM doxorubicin (Sigma-Aldrich, D1515), 1 µM topotecan (Sigma-Aldrich, T2705), 25 µM etoposide (Sigma-Aldrich, E1383). For the inhibitors, cells were incubated with 50 µM aphidicolin (Santa Cruz, sc-201535A), 50 µM Mirin (Selleckchem, S8096), 1 µM VE-821 inhibitor (Selleckchem, S8007), 5 µM KU60019 (Selleckchem, S1570), 400 nM latrunculin B (Sigma-Aldrich, L5288), 5 µM wiskostatin (Sigma-Aldrich, W2270), 50 µM CK666 (Sigma-Aldrich, SML0006) and 50 µM SMIFH2 (Sigma-Aldrich, S4826), as indicated in the figure legends. Investigators were blinded to mouse and cell line genotypes or treatment conditions during experiments and when performing sample analysis, imaging and quantification.

Scratch assay

Scratch assays were performed to evaluate the effect of *Rhoj* KD on the migration of $EPCAM^-$ tumour cells. The cells were grown in a culture-insert

well in 48-well plates. At confluency, cells were serum-starved (1% FBS medium) for 24 h before removing the culture-insert creating a cell-free gap. Next, cells were washed twice with PBS and reincubated in 1% FBS medium. To evaluate wound closure, pictures were processed using ImageJ.

Cell growth assay (crystal violet assay)

For proliferation testing, cells were first seeded in a 96-well plate. After incubation, cells were washed twice with PBS and fixed with 1% glutaraldehyde at different timepoints as described in the figure legends. The staining was then performed using a 0.2% crystal violet solution, and cell were permeabilized using Triton X-100. The absorbance of each well was read at 570 nm on a microplate reader. The cell growth was measured by the ratio of the absorbance of the well at each timepoint to the average absorbance of the wells at day 0.

Tumour transplantation assays

To test the long-term response of *Rhoj*-WT and *Rhoj*-KO $EPCAM^{-}$ tumour cells to chemotherapy, we injected subcutaneously 1,000 cells into NUDE mice after collection in 4 °C medium and resuspension in half medium/half Matrigel (E1270, 970 mg ml^{-1}; Sigma-Aldrich). When the mice developed palpable tumours, they were treated weekly with 3.5 mg per kg cisplatin and 15 mg per kg 5FU as described above. The mice were followed by daily observation and weight measurement, and the tumour size was monitored every week.

Immunoprecipitation

Co-immunoprecipitation of HA-tagged RHOJ was performed using standard methods and adapting previously described protocols[55]. In brief, after three washes in ice-cold Tris-buffered saline, cells were collected on ice by scraping in lysis buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 2 mM EDTA, 10% glycerol, 1 mM β-mercaptoethanol, 1% Triton X-100, protease inhibitor cocktail (11836170001, Roche) and phosphatase inhibitor cocktail 2 (Sigma-Aldrich, P5726)), vortexed 3 times for 30 s with a 2 min pause in between and then centrifuged for 15 min at 13,000 rpm. We evaluated the

total protein content of each sample using the Bradford assay and 1 mg of protein was used in every immunoprecipitation. Antibodies (6 µg; rabbit IgG control Chip grade, ab171870, Abcam; rabbit IPO9, A305-475A, Bethyl Lab) were incubated with 1 mg of precleared lysate at 4 °C under constant rotation overnight. Subsequently, 25 µl Dynabeads Protein G (10003D, Thermo Fisher Scientific) was added and rotated at 4 °C for 4 h. The Dynabeads were washed in 800 µl NETN buffer (20 mM Tris (pH 8), 1 mM EDTA, 900 mM NaCl, 0.5% CA-630) with rotation at 4 °C for 5 min. The washes were repeated five more times and finally the Dynabeads were rinsed with 1 ml dPBS and subsequently eluted with 40 µl 1× SDS gel-loading buffer at 95 °C for 5 min. 26 µl eluted samples were used to perform the western blot using mouse anti-HA antibodies (ab1424, Abcam).

For analyses using liquid chromatography coupled with tandem MS (LC–MS/MS), the lysates were not precleared and the Dynabeads were washed once in wash buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 2 mM EDTA, 10% glycerol, 1 mM β-mercaptoethanol, 0.1% Triton X-100, protease inhibitor cocktail and phosphatase inhibitor cocktail) followed by three washes in MS-compatible buffer (20 mM Tris-HCl pH 8.0, 2 mM $CaCl_2$) before being frozen.

Histology and immunostaining

For immunostaining of frozen sections, skin tumours were embedded in optimal cutting temperature compound (OCT, Sakura) and cut into 5 µm frozen sections using the CM3050S Leica Cryostat (Leica Microsystems) after being pre-fixed in 4% paraformaldehyde for 4 h at room temperature, rinsed in PBS and incubated overnight in 30% sucrose at 4 °C. The sections were blocked using blocking buffer for 1 h (PBS, 5% horse serum, 1% BSA, Triton 0.1%) and then incubated with primary antibodies diluted in blocking buffer overnight at 4 °C, washed three times with PBS for 5 min, and then incubated with Hoechst solution and secondary antibodies diluted in blocking buffer for 1 h at room temperature. Finally, the sections were washed three times with PBS 5 min at room temperature and mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma-Aldrich). Haematoxillin and eosin stainings were performed on paraffin sections, 5 µm paraffin sections were deparaffinized and rehydrated. Slides were

mounted using SafeMount (Labonord). For immunostaining of cultured cells, the cells were plated on a coverslip and fixed with 4% paraformaldehyde for 5 min 48 h after plating the cells. For DDR nuclear focus detection, the coverslips were fixed with 4% paraformaldehyde for 5 min and, after washing, were incubated either in 70% ethanol for 20 min at −20 °C for γ-H2AX and 53bp1 staining or in methanol for 20 min at −20 °C for RAD51 and PCNA staining.

Non-specific antibody binding was prevented by blocking with 5% horse serum, 1% BSA and 0.2% Triton X-100 for 1 h at room temperature. When mouse primary antibodies were used, non-specific antigen blocking was performed using the M.O.M. Basic kit reagent (Vector Laboratories) according to the manufacturer's instructions. Coverslips were then incubated overnight at 4 °C in the presence of the primary antibodies, followed by 1 h of incubation of the secondary antibodies and Hoechst solution at room temperature. Coverslips were mounted using Glycergel (Dako) supplemented with 2.5% DABCO (Sigma-Aldrich). For F-Actin immunofluorescence, the following modifications from the above protocol were made: the coverslips were incubated with blocking solution for 30 min, then incubated for 30 min with rhodamine Phalloidin and, after washes, were incubated with Hoechst solution for 20 min.

Image acquisition and data analysis for immunofluorescence microscopy

Imaging of tumour tissue was performed on the Zeiss Axio Imager M1 (Thornwood) fluorescence microscope with the Zeiss Axiocam MRm camera for immunofluorescence and Zeiss Axiocam MRC5 camera for bright-field microscopy using Axiovision release 4.6 software. Quantification of Ki-67-positive tumour cells was performed on four different fields per tumour using ImageJ. Cell pictures were acquired using the Zeiss Axio Imager.M1 supplied by a ×100 objective (alpha plan-fluar 1.4 numerical aperture oil-immersion objective). *z* series were acquired at 0.3 μm intervals throughout the entire nucleus. Orthogonal projection was performed using Zen Blue 3.3 (ZEISS) to quantify the number of DNA-damage-induced repair foci. The number of foci in each cell was quantified using ImageJ. To determine the nuclear localization of RHOJ, *z* stacks were

processed using the Huygens Professional 22.04 deconvolution software and the global intersection coefficient was used to characterize the degree of overlap between HA-tagged RHOJ and DAPI signals. Brightness, contrast and picture size were adjusted using Photoshop CS6 (Adobe).

Western blot analysis

Total cell lysates were prepared using the radioimmunoprecipitation assay supplemented with phosphatase inhibitors (Cell Signaling, 5870) for 15 min on ice and then centrifuged for 15 min at 14,000 rpm. The protein concentrations of supernatants were measured using the Bradford assay. Total protein lysate (30 µg) was loaded on a gel. Gel electrophoresis was performed using 4–12% NuPAGE Bis-Tris gradient gels (Invitrogen), transferred to PVDF membranes. Membranes were blocked for 1 h in Tris-buffered saline, 0.1% Tween-20 (TBST) containing 3% bovine serum albumin (BSA) then incubated with primary antibodies in the blocking buffer overnight at 4 °C. After washing in TBST, membranes were incubated with secondary antibodies in the blocking buffer for 1 h. Proteins were detected by enhanced chemiluminescence (ECL) western blotting detection reagents (Amersham Biosciences).

FACS isolation of $EPCAM^{+}$ and $EPCAM^{-}$ tumour cells

Tumours were dissected, minced and digested in collagenase I (Sigma-Aldrich) during 2 h at 37 °C on a rocking plate. Collagenase I activity was blocked by the addition of EDTA (5 mM) and the cells were then rinsed in PBS supplemented with 2% FBS. Before the staining, cells were blocked for 20 min at room temperature in PBS supplemented with 30% FBS. Cell suspensions were filtered through a 70 µm cell strainer (BD) then through a 40 µm cell strainer to ensure the elimination of cell debris and clumps of cells. Immunostaining was performed using PE-conjugated anti-CD45 (30F11, 1:100, eBioscience), PE-conjugated anti-CD31 (MEC13.3; 1:100, BD PharMingen) and APC-Cy7-conjugated anti-EPCAM (G8.8; 1:100, BioLegend) for 30 min at 4 °C on a rocking plate protected from light. Living tumour cells were selected by forward scatter, side scatter, doublets discrimination and by Hoechst dye exclusion. $EPCAM^{+}$ and $EPCAM^{-}$ tumour cells were selected on the basis of the expression of YFP and the

exclusion of CD45, CD31 (Lin^-). For $EPCAM^-$ tumour cell subpopulation identification, brilliant violet stain buffer (BD Bioscience) was added (50 µl per sample) and the cells were incubated with PE-conjugated anti-CD51 (rat, RMV-7, BioLegend 104106, 1:50), BV421-conjugated anti-CD61 (Armenian hamster, 2C9.G2, BD Bioscience 553345, 1:50), biotin-conjugated anti-CD106 (rat, 429 (MVCAM.A), BD Bioscience, 553331, 1:50), BV711-conjugated anti-EPCAM (rat, G8.8, BD Bioscience, 563134, 1:100, PerCPCy5.5 conjugated anti-CD45 (rat, 30-F11, BD Bioscience, 550994, 1:100) and PerCPCy5.5 conjugated anti-CD31 (rat, MEC13.3, BD Bioscience, 562861, 1:100) for 30 min at 4 °C. Cells were washed with PBS supplemented with 2% FBS and incubated with streptavidin-BV786 (BD Bioscience, 563858, 1:400) for 30 min at 4 °C. FACS analysis was performed using FACSAria and FACSDiva software (BD Bioscience). Sorted cells were collected either in culture medium for cell culture experiments or into lysis buffer for RNA extraction.

FACS analysis

After trypsinization and washing once in cold PBS, cells were incubated with APC-Cy7-conjugated anti-EPCAM (G8.8; 1:100, BioLegend) antibodies diluted in 200 µl PBS supplemented with 2% FBS for 30 min at 4 °C on a rocking plate protected from light before the fixation step required for the intracellular antigens staining. To detect cell apoptosis or double-stranded DNA breaks, cells were respectively labelled with PE anti-active caspase-3 (BD PharMingen, 550821, 1:25) and PE anti-H2AX (pS139) (BD PharMingen, 562377, 1:20) using the active caspase-3 apoptosis kit (BD PharMingen, 550480) according to the manufacturer's protocol and resuspended in PBS supplemented with 2% FBS. For cell cycle distribution analysis, cells were incubated with 10 µM 5-bromo-2′-deoxyuridine (BrdU) for 45 min. Cells were labelled for BrdU incorporation with an Alexa Fluor 647 anti-BrdU (BD PharMingen, 560209, 1:50) antibody using the BrdU flow kit (BD Pharmigen, 55789) according to the manufacturer's protocol and resuspended in 20 µl 7-AAD for 5 min followed by 200 µl PBS supplemented with 2% FBS. For the chemotherapy sensitivity assay, cells were seeded at an equal density in a six-well plate. Then, 24 h after seeding, the cells were treated with chemotherapy as indicated in the figure legends. Next, 24 h and 48 h after the start of treatment, cells were collected by

trypsinization and quantified by counting the number of living cells by FACS. Living cells were selected by forward and side scatter and by Hoechst dye exclusion. FACS analysis was performed using Fortessa, FACSDiva software and FlowJo (BD Bioscience).

Antibodies

For immunostaining, the following primary antibodies were used: goat GFP (Abcam, ab6673, 1:500), chicken K14 (Thermo Fisher Scientific, MA5-11599, 1:2,000), rabbit vimentin (Abcam, ab92547, 1:500), rabbit active caspase-3 (R&D, AF835, 1:600), rabbit 53bp1 (Novus, NB100-304, 1:200), mouse phospho-histone H2A.X (Ser139) (Millipore, 05-636, 1:500), mouse RAD51 (Santa-Cruz Biotechnology, sc-398587, 1:50), rat RPA32/RPA2 (Cell Signaling, 2208, 1:500), Rhodamine Phalloidin (Thermo Fisher Scientific, Arg415, 1:400), rabbit HA (Abcam, ab9110, 1:1,000), rat phospho-histone H3 (S28) (Abcam, ab10543, 1:2,000), rabbit Ki-67 (Abcam, ab15580, 1:400).

For western blotting, the following primary antibodies were used: rabbit phospho-histone H2A.X (Ser139) (Cell Signaling, 2577, 1:800), rabbit histone H2A.X (Cell Signaling, 2595, 1:1,000), rabbit phosphorylated ATM/ATR substrate (Cell Signaling, 9607, 1:750), rabbit beta-actin (Abcam, ab8227, 1:2,000), mouse HA (Roche, 11583816001, 1:1,000), rabbit phosphorylated CDC2 (Tyr15) (Cell Signaling, 4539, 1:1,000), rabbit phosphorylated CDC2 (Thr161) (Cell Signaling, 9114, 1:1,000), rabbit CDC2 (Cell Signaling, 77055, 1:1,000), rabbit phosphorylated CDK2 (Thr160) (Cell Signaling, 2561, 1:1,000), rabbit CDK2 (Cell Signaling, 18048, 1:1,000), affinity-purified mouse monoclonal antibodies NB8-AD9 (WB/IP) raised against human phosphorylated CDK4 (Thr172)[56] (1:500), rabbit CDK4 (Abcam, ab199728, 1:1,000), rabbit CDK6 (Cell Signaling, 3136, 1:1,000), rabbit N-WASP (Cell Signaling, 4848, 1:1,000), rabbit MCM2 and rabbit MCM3[61], mouse POLD (Santa Cruz, sc-373731), mouse PCNA (Santa Cruz, sc-56), rat RPA32 (Cell Signalling, 2208), rabbit phosphorylated RPA32 S4/S8 (Bethyl, A300-245A), rabbit CTCF (Millipore, 07-729), rabbit H3 (Abcam, ab1791) and mouse α-tubulin (Sigma-Aldrich, T9026). The following secondary antibodies were used:

ECL anti-rabbit, anti-rat anti-mouse IgG conjugated with horseradish peroxidase (GE Healthcare, 1:2,000 or 1:5,000).

Nuclear actin Chromobody transfection

The Nuclear Actin ChromobodyTagGFP plasmid (pnACTagGFP) was purchased from Chromotek. Cells were plated on sterile cover slips and transfected 24 h later with Lipofectamine 3000 (Thermo Fischer Scientific). The chemotherapy agents (cisplatin/5FU) were added 36 h after transfection for 12 and 24 h. After PBS washes, samples were fixed with 4% paraformaldehyde for 5 min at room temperature, counterstained with Hoescht and then mounted with Glycergel (Dako) supplemented with 2.5% DABCO (Sigma-Aldrich). For co-immunostaining of EdU with p-H3 and nuclear actin chromobody–GFP, cells grown onto round coverslips were treated as described in the figures and incubated with 10 μM EdU for the last 45 min. The coverslips were then washed three times with PBS and fixed with 4% paraformaldehyde for 5 min at room temperature. Permeabilization steps and detection of EdU were performed using click-iT Plus EdU cell proliferation kit, Alexa Fluor 647 dye (Thermo Fischer Scientific, C10640), according to the manufacturer's protocol. The coverslips were then stained with p-H3 and Hoescht before proceeding to mounting as previously described.

Virus production, infection and selection

Stable KD cell lines were generated using lentiviral pLKO/PuroR vectors (Sigma-Aldrich) after puromycin selection. KD was confirmed by RT–qPCR. Three different shRNA were used at the same time to target the same gene. Overexpression cell lines were generated using pLX302-EF1a lentiviral vector. pLX302-EF1a was a gift from the Beronja laboratory (Fred Huchinson Cancer Research Center). N-terminal 3×HA-tagged *Rhoj* construct was cloned into pLX302-EF1a using Gateway Technology (Invitrogen) according to the manufacturer's instructions. For virus production, 5×10^6 HEK293T cells were seeded into 10 cm dishes and transfected with the vector of interest and appropriate packaging plasmids psPax2 and pMD2.G (12260 and 12259, respectively, Addgene). The medium was changed 24 h later and supernatants were next collected at 48 h,

and passed through a 0.45 μm filter. Tumour epithelial cells or tumour mesenchymal cells were plated in six-well plates and incubated with 40 μl per ml viruses when they reach 50% of confluence, in the presence of polybrene (5 mg ml^{-1}). Medium was changed 24 h later and cells were selected with puromycin for at least 1 week.

shRNA

The following shRNA were used (TRC ID, clone name): mRhoj: TRCN0000313500, NM_023275.2, 673s21c1; mRhoj: TRCN0000313499, NM_023275.2, 713s21c1; mRhoj: TRCN0000077567, NM_023275.1, 690s1c1; mRhoq: TRCN0000077513, NM_145491.2, 1732s1c1; mRhoq: TRCN0000077515, NM_145491.2, 892s1c1; mRhoq: TRCN0000077516, NM_145491.2, 661s1c1; hRHOJ: TRCN0000047603, NM_020663.2, 878s1c1; hRHOJ: TRCN0000047604, NM_020663.2, 1007s1c1; hRHOJ: TRCN0000047605, NM_020663.2, 647s1c1.

RNA and DNA extraction and RT–qPCR

RNA extraction from FACS-isolated cells was performed using the RNeasy micro kit (QIAGEN) according to the manufacturer's recommendations with DNase treatment. After RNA quantification using the Nanodrop, the first-strand cDNA was synthesized using Superscript II (Invitrogen) and random hexamers (Roche) at a final volume of 50 μl. Control of genomic contaminations was measured for each sample by performing the same procedure with or without reverse transcriptase. qPCR assays were performed using 1 ng of cDNA as template, SYBRGreen mix (Applied Bioscience) and the Light Cycler 96 (Roche) real-time PCR system. The *Tbp* housekeeping gene was used for normalization of mouse RT–qPCR and the *POLR2A* housekeeping gene was used for normalization of human RT–qPCR. Primers were designed using Roche Universal ProbeLibrary Assay Design Center (https://lifescience.roche.com/webapp/wcs/stores/servlet/CategoryDisplay?tab=Assay+Design+Center&identifier=Universal+Probe+Library&langId=-1). qPCR analysis was performed using the Light Cycler 96 (Roche) and the $\Delta\Delta C_t$ method with *Tbp* as a reference.

Primers used for RT–qPCR

The sequences of the primers used for RT–qPCR were as follows: mRhoj For(5′-3′): ACCACTACGCAGTTACCGTG; mRhoj Rev(3′-5′): TGCAACACCATTCTCCGACC; mRhoq For(5′-3′): TTCGACCACTACGCAGTCAG; mRhoq Rev(3′-5′): CCTGCAAACCGCGTATAAGG; mTBP For(5′-3′): TGTACCGCAGCTTCAAAATATTGTAT; mTBP Rev(3′-5′): AAATCAACGCAGTTGTCCGTG; mKrt14 For(5′-3′): GCCGCCCCTGGTGTGGAC; mKrt14 Rev(3′-5′): GTGCGCCGGAGCTCAGAAATC; mEpcam For(5′-3′): CATTTGCTCCAAACTGGCGT; mEpcam Rev(3′-5′): TTGTTCTGGATCGCCCCTTC; mVimFor(5′-3′): CCAACCTTTTCTTCCCTGAAC; mVimRev(3′-5′): TTGAGTGGGTGTCAACCAGA; mZeb1 For(5′-3′): ATTCCCCAAGTGGCATATACA; mZeb1 Rev(3′-5′): GAGCTAGTGTCTTGTCTTTCATCC; mZeb2 For(5′-3′): ACCGCATATGGCCTATACCTAC; mZeb2 Rev(3′-5′): TGCTCCATCCAGCAAGTCT; mPrrx1 For(5′-3′): TGTTGATTCGAGCGGGAAGA; mPrrx1 Rev(3′-5′): TCTAGCAGGTGACTGACGGA; mPdgfra For(5′-3′): TGGAAGCTTGGGGCTTACTT; mPdgfra Rev(3′-5′): CATAGCTCCTGAGACCTTCTCC; hRHOJ For(5′-3′): ACAATGTCCAGGAGGAATGGG; hRHOJ Rev(3′-5′): TGTGCTCCGATCGCTTTTG; hRHOQ For (5′-3′): AAAGAGGAGTGGGTACCGGA; hRHOQ Rev (5′-3′): GCAGCATGCTCCTATCTCTT; hPOLR2A For(5′-3′): GCAAATTCACCAAGAGAGACG; and hPOLR2A Rev(3′-5′): CACGTCGACAGGAACATCAG.

RNA-seq analysis

RNA quality was checked using the Bioanalyzer (Agilent). For RNA extracted from tumour epithelial cells or tumour mesenchymal cells, indexed cDNA libraries were obtained using the TruSeq Stranded mRNA Sample prep kit (Illumina) according to the manufacturer’s recommendations. The multiplexed libraries (11 pM) were loaded and sequences were produced

using a HiSeq PE Cluster Kit v4 and TruSeq SBS Kit v3-HS (250 cycles) on the HiSeq 1500 (Illumina) system. Approximately 8 million paired-end reads per sample were mapped against the mouse reference genome (GRCm38.p4/mm10) using STAR software to generate read alignments for each sample. Annotations Mus_musculus.GRCm38.84.gtf were obtained from https://ftp.ensembl.org/. After transcripts were assembled, gene-level counts were obtained using HTSeq and normalized to 20 million aligned reads. Fold change (FC) values were computed on these values between the conditions. The accession number for the RNA-seq reported in this paper is GEO: GSE205985.

DNA fibre analysis

Exponentially growing cells were pulse-labelled with 50 µM CldU (20 min; Sigma-Aldrich, C6891) followed by 250 µM IdU (20 min; Sigma-Aldrich, I7125). Labelled cells were collected and DNA fibres were spread in buffer containing 0.5% SDS, 200 mM Tris pH 7.4 and 50 mM EDTA as described previously[57]. For immunodetection of labelled tracks, fibres were incubated with primary antibodies (for CldU, rat anti-BrdU, ab6326 Abcam; for IdU, mouse anti-BrdU, 347580 BD Bioscience) for 1 h at room temperature and developed with the corresponding secondary antibodies (anti-rat IgG AF594, A-11007; anti-mouse IgG1 AF488, A-21121; all from Molecular Probes) for 30 min at room temperature. Mouse anti-ssDNA antibody was used to assess fibre integrity (MAB3034, Millipore, secondary antibody anti-mouse IgG2a AF647, A-21241 Molecular Probes). Slides were examined with a Leica DM6000 B microscope, as described previously[58]. The conversion factor used was 1 µm = 2.59 kb (ref. [59]). In each assay, more than 250 tracks were measured to estimate the fork rate and more than 500 tracks were analysed to estimate the frequency of origin firing (first label origins—green-red-green—are shown as the percentage of all red (CldU-labelled) tracks)[60]. For estimating asymmetry, both forks emanating from an origin were measured (100 origins per condition in each experiment) and the ratio of long/short was calculated. An origin was considered to be asymmetric when the long/short fork ratio was greater than 1.4.

Chromatin fractionation

For the analysis of chromatin-bound proteins, biochemical fractionations were performed as described previously[61]. In brief, cells were resuspended at 10^7 cells per ml in buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 1× protease inhibitor cocktail; bimake.com), and incubated on ice for 5 min in the presence of 0.1% Triton X-100. Low-speed centrifugation (for 4 min at 600*g* and 4 °C) enabled the separation of the cytosolic fraction (supernatant) and nuclei (pellet). Nuclei were washed and subjected to hypotonic lysis in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, 1× protease inhibitor cocktail) for 30 min on ice. The nucleoplasmic and chromatin fractions were separated after centrifugation (for 4 min at 600*g* and 4 °C). Chromatin was resuspended in Laemmli sample buffer. Whole-cell extracts (WCE) were prepared by resuspension of cells in Laemli sample buffer (10^7 cells per ml). Both WCE and chromatin were sonicated twice for 15 s at 15% amplitude and boiled at 95 °C for 5 min. Then, 10 µl of WCE (equivalent to 10^5 cells) and 20 µl of chromatin (2×10^5 cells) were loaded onto 4–20% custom-made gradient gels. Protein was transferred to nitrocellulose membranes, blocked with 5% non-fat milk in TBST for 1 h at room temperature and incubated with primary antibodies overnight at 4 °C. Incubation with secondary antibodies was performed for 1 h at room temperature and proteins were detected with WesternBright ECL (Advansa).

EdU immunofluorescence

Cells grown onto round coverslips were treated as described in the figure and incubated with 25 µM EdU for the last 30 min. The coverslips were then washed twice with PBS and cells were pre-extracted with CSK buffer (10 mM PIPES pH 7, 0.1 M NaCl, 0.3 M sucrose, 3 mM $MgCl_2$) for 5 min 4 °C. After fixation with formaldehyde for 10 min at room temperature, a click reaction was performed on the coverslips (100 mM Tris pH 8, 10 mM $CuSO_4$, 2 mM Na-l-ascorbate, 50 µM biotin-azide-AF488) for 1.5 h at 37 °C. Then, the coverslips were washed with PBS and DNA was stained with DAPI (1 mg ml^{-1} in PBS) for 10 min at room temperature, mounted with Prolong Gold Antifade (Thermo Fisher Scientific) and visualized under a fluorescence microscope (DM6000 B Leica microscope with a HCX PL APO 40 0.75 NA objective). EdU intensity was assessed in at least 250

EdU-positive nuclei per condition per experiment using CellProfiler (v.3.1.9).

Sample preparation for proteomic analysis

To define the effect of RHOJ on protein expression in EMT tumour cells, a total of 24 samples was prepared for LC–MS/MS analyses, corresponding to 3 replicates of $EPCAM^{+}$ tumour cells, 3 replicates of $EPCAM^{-}$ tumour cells, 3 replicates of $EPCAM^{-}$ *Rhoj* shRNA tumour cells and 3 replicates of $EPCAM^{-}$ control shRNA treated for 24 h with cisplatin/5FU as described above or with vehicle control (DMSO). Cells were collected by manual scraping in Tris-buffered saline on ice and were flash-frozen as a dry pellet. The cell pellets (15 million cells per pellet) were lysed in 1 ml urea lysis buffer containing 9 M urea, 20 mM HEPES pH 8.0 and PhosSTOP phosphatase inhibitor cocktail (Roche, 1 tablet per 10 ml buffer). The samples were sonicated with 3 pulses of 15 s at an amplitude of 20% using a 3 mm probe, with incubation on ice for 1 min between pulses. After centrifugation for 15 min at 20,000*g* at room temperature to remove insoluble components, proteins were reduced by addition of 5 mM DTT and incubation for 30 min at 55° C and then alkylated by addition of 10 mM iodoacetamide and incubation for 15 min at room temperature in the dark. The protein concentration was measured using the Bradford assay (Bio-Rad) and, from each sample, 1 mg protein was used to continue the protocol. The samples were further diluted with 20 mM HEPES pH 8.0 to a final urea concentration of 4 M and proteins were digested with 10 µg LysC (Wako) (1/100, w/w) for 4 h at 37 °C. Samples were again diluted to 2 M urea and digested with 10 µg trypsin (Promega) (1/100, w/w) overnight at 37° C. The resulting peptide mixture was acidified by addition of 1% trifluoroacetic acid and, after 15 min incubation on ice, the samples were centrifuged for 15 min at 1,780*g* at room temperature to remove insoluble components. Next, peptides were purified on SampliQ SPE C18 cartridges (Agilent). Columns were first washed with 1 ml 100% acetonitrile and pre-equilibrated with 3 ml of solvent A (0.1% trifluoroacetic acid in water/acetonitrile (98:2, v/v)) before the samples were loaded on the column. After peptide binding, the column was washed again with 2 ml of solvent A and peptides were eluted twice with 750 µl elution buffer (0.1% trifluoroacetic acid in

water/acetonitrile (20:80, v/v)). Then, 75 µl of the eluate was dried completely in a SpeedVac vacuum concentrator for shotgun analysis.

LC–MS/MS analysis

Purified peptides were redissolved in 30 µl solvent A and half of each sample was injected for LC–MS/MS analysis on the Ultimate 3000 RSLCnano system in-line connected to a Q Exactive HF mass spectrometer equipped with a Nanospray Flex Ion source (Thermo Fisher Scientific). Trapping was performed at 10 µl min^{-1} for 4 min in loading solvent A on a 20 mm trapping column (custom made, 100 µm internal diameter, 5 µm beads, C18 Reprosil-HD, Dr Maisch) and the sample was loaded onto a reverse-phase column (custom made, 75 µm inner diameter × 400 mm, 1.9 µm beads C18 Reprosil-HD, Dr Maisch). The peptides were eluted by a nonlinear increase from 2% to 56% solvent B (0.1% formic acid in water/acetonitrile (20:80, v/v)) over 145 min at a constant flow rate of 250 nl min^{-1}, followed by a 10 min wash reaching 99% solvent B and re-equilibration with solvent A (0.1% formic acid in water). The column temperature was kept constant at 50 °C (CoControl v.3.3.05, Sonation).

The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the 16 most abundant ion peaks per MS spectrum. Full-scan MS spectra (375 to 1,500 *m*/*z*) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target value of 3,000,000. The 16 most intense ions above a threshold value of 13,000 were isolated (window of 1.5 Th) for fragmentation at a normalized collision energy of 28% after filling the trap at a target value of 100,000 for maximum 80 ms. MS/MS spectra (200 to 2,000 *m*/*z*) were acquired at a resolution of 15,000 in the orbitrap analyzer. The S-lens RF level was set at 55, and we excluded precursor ions with single and unassigned charge states from fragmentation selection. QCloud[62] was used to control instrument longitudinal performance during the project.

Proteomic data analysis

Data analysis of the shotgun data was performed using MaxQuant (v.1.5.8.3) using the Andromeda search engine with the default search settings

including a false-discovery rate set at 1% at the peptide and protein level. Spectra were searched against the mouse proteins in the Swiss-Prot database (database release version of September 2017 containing 16,931 mouse protein sequences; http://www.uniprot.org). The mass tolerance for precursor and fragment ions was set to 4.5 and 20 ppm, respectively, during the main search. Enzyme specificity was set as C terminal to arginine and lysine, also allowing cleavage at proline bonds with a maximum of two missed cleavages. Variable modifications were set to oxidation of methionine residues, acetylation of protein N termini, and phosphorylation of serine, threonine or tyrosine residues, while carbamidomethylation of cysteine residues was set as a fixed modification. Matching between runs was enabled with a matching time window of 0.7 min and an alignment time window of 20 min. Only proteins with at least one unique or razor peptide were retained leading to the identification of 5,354 proteins. Proteins were quantified by the MaxLFQ algorithm integrated in the MaxQuant software. A minimum ratio count of two unique or razor peptides was required for quantification.

Further data analysis of the shotgun results was performed using the Perseus software (v.1.5.5.3) after loading the protein groups file from MaxQuant. Reverse database hits, potential contaminants and hits only identified by site were removed, LFQ intensities were $\log_2$-transformed and replicate samples were grouped. Proteins with less than three valid values in at least one group were removed and missing values were imputed from a normal distribution around the detection limit leading to a list of 4,239 quantified proteins that was used for further data analysis. To reveal proteins of which the expression level was significantly regulated between the different conditions, sample groups were defined on the basis of the treatment (control versus treated) and RHOJ expression (with versus without RHOJ) and a two-way ANOVA test was performed. For each protein, this test calculated a P value (−log-transformed P value) for treatment, a P value for RHOJ expression and a P value for the interaction between treatment and RHOJ expression. Proteins with $P < 0.05$ in at least one of these three conditions were considered to be significantly regulated. A total of 99 proteins of interest were selected, the $\log_2$ transformed intensities of these proteins were Z-scored and these values were plotted in a heat map after non-supervised hierarchical clustering.

The MS proteomics data have been deposited at the ProteomeXchange Consortium through the PRIDE[63] partner repository under dataset identifier PXD025737.

Sample preparation for AP–MS

To identify the proteins interacting with RHOJ, a total of 12 samples were prepared for LC–MS/MS analysis corresponding to 3 replicates of HA-tagged-RHOJ-transfected EPCAM^{-} tumour cells and 3 control empty vector EPCAM^{-} tumour cells treated for 12 h with cisplatin/5FU or with vehicle control (DMSO). Immunoprecipitation of HA-tagged RHOJ was performed as follows. After three washes in ice-cold Tris-buffered saline, cells were collected on ice by scraping in lysis buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 2 mM EDTA, 10% glycerol, 1 mM β-mercaptoethanol, 1% Triton X-100, protease inhibitor cocktail (11836170001, Roche) and phosphatase inhibitor cocktail 2 (Sigma-Aldrich, P5726)), vortexed three times for 30 s with a 2 min pause in between and then centrifuged for 15 min at 13,000 rpm. The total protein content of each sample was evaluated using the Bradford assay and 1 mg of protein was used in every immunoprecipitation. A total of 6 µg of antibodies (rabbit HA tag Chip grade, ab9110, Abcam) were incubated with 1 mg of lysate at 4 °C under constant rotation overnight. Subsequently, 25 µl Dynabeads Protein G (10003D, Thermo Fisher Scientific) was added and rotated at 4 °C for 4 h. The Dynabeads were washed once in wash buffer (150 mM NaCl, 50 mM HEPES pH 7.5, 2 mM EDTA, 10% glycerol, 1 mM β-mercaptoethanol, 0.1% Triton X-100, protease inhibitor cocktail and phosphatase inhibitor cocktail) followed by three washes in MS-compatible buffer (20 mM Tris-HCl pH 8.0, 2 mM $CaCl_2$). Washed beads were resuspended in 150 µl trypsin digestion buffer and incubated for 4 h with 1 µg trypsin (Promega) at 37° C. Beads were removed, another 1 µg of trypsin was added and proteins were further digested overnight at 37° C. Peptides were purified on Omix C18 tips (Agilent) and dried completely in a rotary evaporator.

LC–MS/MS analysis

Peptides were redissolved in 20 µl loading solvent A (0.1% trifluoroacetic acid in water/acetonitrile (98:2, v/v)) of which 2 µl was injected for LC–

MS/MS analysis on the Ultimate 3000 RSLCnano system in-line connected to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). Trapping was performed at 10 µl min^{-1} for 2 min in loading solvent A on a 5 mm trapping column (Thermo Fisher Scientific, 300 µm internal diameter, 5 µm beads). The peptides were separated on a 250 mm Waters nanoEase M/Z HSS T3 Column, 100 Å, 1.8 µm, 75 µm inner diameter (Waters Corporation) kept at a constant temperature of 45 °C. Peptides were eluted by a non-linear gradient starting at 1% MS solvent B reaching 33% MS solvent B (0.1% formic acid in water/acetonitrile (2:8, v/v)) in 60 min, 55% MS solvent B (0.1% formic acid in water/acetonitrile (2:8, v/v)) in 75 min, 99% MS solvent B in 90 min followed by a 10 min wash at 99% MS solvent B and re-equilibration with MS solvent A (0.1% formic acid in water). The mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the 12 most abundant ion peaks per MS spectrum. Full-scan MS spectra (375–1,500 *m*/*z*) were acquired at a resolution of 60,000 in the Orbitrap analyzer after accumulation to a target value of 3,000,000. The 12 most intense ions above a threshold value of 15,000 were isolated with a width of 1.5 *m*/*z* for fragmentation at a normalized collision energy of 30% after filling the trap at a target value of 100,000 for maximum 80 ms. MS/MS spectra (200–2,000 *m*/*z*) were acquired at a resolution of 15,000 in the Orbitrap analyzer.

Data analysis

Analysis of the MS data was performed in MaxQuant (v.2.0.3.0) with mainly the default search settings, including a false-discovery rate set at 1% at the peptide-to-spectrum match, peptide and protein level. Spectra were searched against the mouse proteins in the Reference proteins database (UP000000589, database release version of January 2022 containing 21,986 mouse protein sequences; http://www.uniprot.org). The mass tolerance for precursor and fragment ions was set to 4.5 and 20 ppm, respectively, during the main search. Enzyme specificity was set as C terminal to arginine and lysine, also allowing cleavage at proline bonds with a maximum of two missed cleavages. Variable modifications were set to oxidation of methionine residues, acetylation of protein N termini. Matching between runs was enabled with a matching time window of 0.7 min and an alignment time window of 20 min. Only proteins with at least one unique or razor

peptide were retained. Proteins were quantified using the MaxLFQ algorithm integrated in the MaxQuant software. A minimum ratio count of two unique or razor peptides was required for quantification. A total of 126,976 peptide-to-spectrum matches was performed, resulting in 14,501 identified unique peptides, corresponding to 2,091 identified proteins. Further data analysis of the AP–MS results was performed using a custom R script, using the proteinGroups output table from MaxQuant. Reverse database hits were removed, LFQ intensities were log_2-transformed and the replicate samples were grouped. Proteins with less than three valid values in at least one group were removed and missing values were imputed from a normal distribution centred around the detection limit (package DEP[64]), leading to a list of 1,381 quantified proteins in the experiment, used for further data analysis. To compare protein abundance between pairs of sample groups (RhoJHAuntreatedIP versus EVuntreatedIP, RhoJHAtreated12hchemoIP versus EVtreated12hchemoIP sample groups), statistical testing for differences between two group means was performed, using the package limma[65]. Statistical significance for differential regulation was set to a false-discovery rate of <0.05 and a fold change of >2-fold or <0.5-fold ($|log_2FC| = 1$). The MS proteomics data have been deposited at the ProteomeXchange Consortium through the PRIDE partner repository under dataset identifier PXD038278.

Statistics

Statistical and graphical data analyses were performed using IBM-SPSS v.28.0 (IBM, Released 2021; IBM SPSS Statistics for Windows, v.28.0), Medcalc v.20 (MedCalc Statistical Software v.20.109; MedCalc Software; https://www.medcalc.org; 2022) and Prism 5 (GraphPad). Continuous variables are summarized by their means and their s.e.m., and qualitative variables as numbers and percentages. Differences in continuous variables were compared between groups using nonparametric tests in the case of small sample sizes (*n* of less than around 30), including Mann–Whitney tests in the case of two groups and Kruskal–Wallis tests followed by Mann–Whitney tests corrected for multiple comparisons with Bonferroni correction when there were more than two groups. When sample sizes were sufficiently large (*n* of more than around 30), parametric tests were used, including classical Student's *t*-tests or Welch's *t*-tests in the case of variance inequality

when there were two samples and ANOVA followed by Sidak test or Dunnett T3 test in the case of variance heterogeneity when more than two samples had to be compared. $P < 0.05$ was considered to be statistically significant. All statistical tests were two-sided. Each test used is mentioned in the legend of the respective figure.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The MS proteomics data have been deposited to the ProteomeXchange Consortium through the PRIDE[63] partner repository under dataset identifiers PXD025737 and PXD038278. RNA-seq data reported in this paper are available at the GEO (GSE205985). All other data are available from the corresponding author on reasonable request. Source data are provided with this paper.

References

1. Zaorsky, N. G. et al. Causes of death among cancer patients. *Ann. Oncol.* **28**, 400–407 (2017).

2. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat. Rev. Cancer* **18**, 128–134 (2018).

3. Shibue, T. & Weinberg, R. A. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat. Rev. Clin. Oncol.* **14**, 611–629 (2017).

4. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).

5. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat. Cell Biol.* **16**, 488–494 (2014).

6. Li, Q. Q. et al. Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells. *Clin. Cancer Res.* **15**, 2657–2665 (2009).

7. Zhang, P. et al. ATM-mediated stabilization of ZEB1 promotes DNA damage response and radioresistance through CHK1. *Nat. Cell Biol.* **16**, 864–875 (2014).

8. Zheng, X. et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).

9. Fischer, K. R. et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).

10. Latil, M. et al. Cell-type-specific chromatin states differentially prime squamous cell carcinoma tumor-initiating cells for epithelial to mesenchymal transition. *Cell Stem Cell* **20**, 191–204 (2017).

11. Schmults, C. D. et al. NCCN Guidelines® insights: squamous cell skin cancer, Version 1.2022. *J. Natl Compr. Canc. Netw.* **19**, 1382–1394 (2021).

12. Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).

13. Ho, H. et al. RhoJ regulates melanoma chemoresistance by suppressing pathways that sense DNA damage. *Cancer Res.* **72**, 5516–5528 (2012).

14. Ruiz, R. et al. The RhoJ-BAD signaling network: an Achilles' heel for BRAF mutant melanomas. *PLoS Genet.* **13**, e1006913 (2017).

15. Chen, B. et al. MKL1 mediates TGF-β induced RhoJ transcription to promote breast cancer cell migration and invasion. *Front Cell Dev. Biol.* **8**, 832 (2020).

16. Kaur, S. et al. RhoJ/TCL regulates endothelial motility and tube formation and modulates actomyosin contractility and focal adhesion numbers. *Arterioscler. Thromb. Vasc. Biol.* **31**, 657–664 (2011).

17. Wang, M. et al. Rhoj is a novel target for progression and invasion of glioblastoma by impairing cytoskeleton dynamics. *Neurotherapeutics* **17**, 2028–2040 (2020).

18. Polo, S. E. & Jackson, S. P. Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes Dev.* **25**, 409–433 (2011).

19. Vignal, E. et al. Characterization of TCL, a new GTPase of the rho family related to TC10 and Cdc42. *J. Biol. Chem.* **275**, 36457–36464 (2000).

20. Higgs, M. R. et al. BOD1L is required to suppress deleterious resection of stressed replication forks. *Mol. Cell* **59**, 462–477 (2015).

21. Bagci, H. et al. Mapping the proximity interaction network of the Rho-family GTPases reveals signalling pathways and regulatory mechanisms. *Nat. Cell Biol.* **22**, 120–134 (2020).

22. Blackford, A. N. & Jackson, S. P. ATM, ATR, and DNA-PK: the trinity at the heart of the DNA damage response. *Mol. Cell* **66**, 801–817 (2017).

23. Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. p53 binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).

24. Chapman, J. R. et al. RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Mol. Cell* **49**, 858–871 (2013).

25. Liu, J. S., Kuo, S. R. & Melendy, T. DNA damage-induced RPA focalization is independent of gamma-H2AX and RPA hyper-phosphorylation. *J. Cell. Biochem.* **99**, 1452–1462 (2006).

26. Murai, J. et al. SLFN11 blocks stressed replication forks independently of ATR. *Mol. Cell* **69**, 371–384 (2018).

27. Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol. Cell* **37**, 492–502 (2010).

28. Cybulla, E. & Vindigni, A. Leveraging the replication stress response to optimize cancer therapy. *Nat. Rev. Cancer* **23**, 6–24 (2023).

29. Cimprich, K. A. & Cortez, D. ATR: an essential regulator of genome integrity. *Nat. Rev. Mol. Cell Biol.* **9**, 616–627 (2008).

30. Byun, T. S., Pacek, M., Yee, M. C., Walter, J. C. & Cimprich, K. A. Functional uncoupling of MCM helicase and DNA polymerase activities activates the ATR-dependent checkpoint. *Genes Dev.* **19**, 1040–1052 (2005).

31. Tye, S., Ronson, G. E. & Morris, J. R. A fork in the road: where homologous recombination and stalled replication fork protection part ways. *Semin. Cell Dev. Biol.* **113**, 14–26 (2020).

32. Técher, H. et al. Replication dynamics: biases and robustness of DNA fiber analysis. *J. Mol. Biol.* **425**, 4845–4855 (2013).

33. Ge, X. Q., Jackson, D. A. & Blow, J. J. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev.* **21**, 3331–3341 (2007).

34. Ibarra, A., Schwob, E. & Méndez, J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc. Natl Acad. Sci. USA* **105**, 8956–8961 (2008).

35. Dopie, J., Skarp, K. P., Rajakylä, E. K., Tanhuanpää, K. & Vartiainen, M. K. Active maintenance of nuclear actin by importin 9 supports transcription. *Proc. Natl Acad. Sci. USA* **109**, E544–E552 (2012).

36. Andrin, C. et al. A requirement for polymerized actin in DNA double-strand break repair. *Nucleus* **3**, 384–395 (2012).

37. Belin, B. J., Lee, T. & Mullins, R. D. DNA damage induces nuclear actin filament assembly by formin-2 and spire-1/2 that promotes efficient DNA repair. *eLlife* **4**, e07735 (2015).

38. Caridi, C. P. et al. Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* **559**, 54–60 (2018).

39. Lamm, N. et al. Nuclear F-actin counteracts nuclear deformation and promotes fork repair during replication stress. *Nat. Cell Biol.* **22**, 1460–1470 (2020).

40. Parisis, N. et al. Initiation of DNA replication requires actin dynamics and formin activity. *EMBO J.* **36**, 3212–3231 (2017).

41. Schrank, B. R. et al. Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61–66 (2018).

42. Lionarons, D. A. et al. $RAC1^{P29S}$ induces a mesenchymal phenotypic switch via serum response factor to promote melanoma development and therapy resistance. *Cancer Cell* **36**, 68–83 (2019).

43. Magalhaes, Y. T., Silva, G. E. T., Osaki, J. H., Rocha, C. R. R. & Forti, F. L. RHOAming through the nucleotide excision repair pathway as a mechanism of cellular response against the effects of UV radiation. *Front. Cell Dev. Biol.* **8**, 816 (2020).

44. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15**, 178–196 (2014).

45. Ho, H. et al. RhoJ modulates melanoma invasion by altering actin cytoskeletal dynamics. *Pigment Cell Melanoma Res.* **26**, 218–225 (2013).

46. Abe, T., Kato, M., Miki, H., Takenawa, T. & Endo, T. Small GTPase Tc10 and its homologue RhoT induce N-WASP-mediated long process

formation and neurite outgrowth. *J. Cell Sci.* **116**, 155–168 (2003).

47. Aymard, F. et al. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat. Struct. Mol. Biol.* **24**, 353–361 (2017).

48. Srinivas, S. et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).

49. Madisen, L. et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

50. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).

51. Tuveson, D. A. et al. Endogenous oncogenic *K-ras*$^{\mathrm{G12D}}$ stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell* **5**, 375–387 (2004).

52. Lapouge, G. et al. Skin squamous cell carcinoma propagating cells increase with tumour progression and invasiveness. *EMBO J.* **31**, 4563–4575 (2012).

53. Kim, C. et al. Vascular RhoJ is an effective and selective target for tumor angiogenesis and vascular disruption. *Cancer Cell* **25**, 102–117 (2014).

54. Lapouge, G. et al. Identifying the cellular origin of squamous skin tumors. *Proc. Natl Acad. Sci. USA* **108**, 7431–7436 (2011).

55. Sambrook, J. & Russell, D. W. Identification of associated proteins by coimmunoprecipitation. *CSH Protoc.* **2006**, pdb.prot3898 (2006).

56. Coulonval, K. et al. Monoclonal antibodies to activated CDK4: use to investigate normal and cancerous cell cycle regulation and involvement of phosphorylations of p21 and p27. *Cell Cycle* **21**, 12–32 (2022).

57. Rodriguez-Acebes, S., Mourón, S. & Méndez, J. Uncoupling fork speed and origin activity to identify the primary cause of replicative stress phenotypes. *J. Biol. Chem.* **293**, 12855–12861 (2018).

58. Mourón, S. et al. Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nat. Struct. Mol. Biol.* **20**, 1383–1389 (2013).

59. Jackson, D. A. & Pombo, A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J. Cell Biol.* **140**, 1285–1295 (1998).

60. Petermann, E., Woodcock, M. & Helleday, T. Chk1 promotes replication fork progression by controlling replication initiation. *Proc. Natl Acad. Sci. USA* **107**, 16090–16095 (2010).

61. Méndez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).

62. Chiva, C. et al. QCloud: a cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PLoS ONE* **13**, e0189209 (2018).

63. Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450 (2019).

64. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).

65. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

Acknowledgements

We thank G. Lapouge and the staff at the ULB animal facility; the staff at the ULB genomic core facility (F. Libert and A. Lefort); M. Martens and J.-M. Vanderwinden and the members of the LiMiF (www.limif.ulb.ac.be) for help with microscopy; B. Beck, O. Serçin, A. Karambelas and Blanpain laboratory members for discussions and suggestions. We thank Y. Pommier for helpful insights. M.D. was supported by FNRS doctoral fellowship (Aspirant) and Association Vésale. M.Z. and M.-A.P. were supported by FNRS fellowship and the TELEVIE. J.M. is supported by the Ministry of Science and Innovation, Spain. The Center for Microscopy and Molecular Imaging (CMMI) is supported by the European Regional Development Fund and the Walloon Region. Work in the Mendez laboratory was supported by grant PID2019-106707-RB from the Spanish Ministry of Science and Innovation, co-sponsored by ERDF funds from the EU. C.B. is supported by WELBIO, FNRS, the TELEVIE, Julie and Françoise Drion Fondation, Fondation Contre le Cancer, ULB Foundation, Foundation Baillet Latour, FNRS/FWO EOS (40007513) and the European Research Council (AdvGrant 885093).

Author information

Authors and Affiliations

1. Laboratory of Stem Cells and Cancer, Université Libre de Buxelles (ULB), Brussels, Belgium

   Maud Debaugnies, Jeremy Blondeau, Marie-Astrid Parent, Manuel Zocco, Yura Song, Virginie Moers, Mathilde Latil, Christine Dubois, Panagiota A. Sotiropoulou & Cédric Blanpain

2. CHU Saint-Pierre, Université Libre de Bruxelles (ULB), Brussels, Belgium

   Maud Debaugnies

3. DNA Replication Group, Molecular Oncology Programme, Spanish National Cancer Research Centre, Madrid, Spain

Sara Rodríguez-Acebes & Juan Méndez

4. Institute of Interdisciplinary Research (IRIBHM), Université Libre de Bruxelles (ULB), Brussels, Belgium

Viviane de Maertelaer & Katia Coulonval

5. ULB-Cancer Research Center (U-crc), Université Libre de Bruxelles (ULB), Brussels, Belgium

Viviane de Maertelaer & Katia Coulonval

6. VIB Center for Medical Biotechnology, VIB Proteomics Core, Department of Biomolecular Medicine, Ghent University, Ghent, Belgium

Francis Impens, Delphi Van Haver & Sara Dufour

7. Department of Retinal Vascular Biology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

Akiyoshi Uemura

8. WELBIO, Université Libre de Bruxelles (ULB), Brussels, Belgium

Cédric Blanpain

Contributions

M.D. and C.B. designed the experiments, performed most of the data analysis and wrote the manuscript. M.D., M.-A.P. and J.B. performed most of the biological experiments. M.Z. contributed to performing gain-of-function experiments and sample collection for proteomic analysis. M.L. contributed to performing shRNA gene KD experiments. S.R.-A. and J.M. conducted DNA fibre analysis, identification of replication factors on the chromatin fraction, EdU incorporation analysis and western blot validation of proteins identified by proteomic analysis, and acquired, analysed and interpreted the data. K.C. performed the western blot to study the activation of different CDKs. Y.S. performed bioinformatic analysis. V.d.M. performed

the statistical analysis. C.D. performed FACS. V.M. provided technical support. F.I., D.V.H. and S.D. performed proteomic analysis. A.U. generated *Rhoj*-cKO mice. P.A.S. contributed to the research design and supervised experiments. All of the authors read and approved the final manuscript.

Corresponding author

Correspondence to Cédric Blanpain.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Stephanie Panier and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Mouse skin SCCs with EMT are resistant to chemotherapy.

a, Scheme of the genetic strategy to induce *Kras*^{G12D} and YFP expression and *Trp53* deletion in the mouse model of skin SCCs. After tumour

appearance, mice are treated for 4 weeks with cisplatin/5FU. **b**, Evolution of tumour size of treated primary SCC. Relative tumour volumes (Mean ± S.D.) are shown for not responding (n = 18), partially responding (n = 30) and fully responding (n = 8) tumours after 4 weeks of cisplatin/5FU treatment. **c**, Proportion of not responding, partially responding and fully responding tumours after 4 weeks of cisplatin/5FU treatment. **d**, Coimmunostaining of YFP and Keratin 14 (K14) or Vimentin (Vim) in partially responding and non-responding tumours treated for 4 weeks with cisplatin/5FU treatment (n = 5; scale bar, 20μm). **e-i**, FACS plots showing the gating strategy used to FACS-isolate or to analyse the proportion of YFP+ EPCAM^{+} and EPCAM^{-} tumour cells from *Lgr5^{creER}Kras^{G12D}Trp53^{cKO}Rosa-YFP* and *Lgr5^{creER}Kras^{G12D}Trp53^{cKO}Rhoj^{cKO}Rosa-YFP* skin SCCs presented on Fig. 2h and Extended Data Fig. 1j. Cells were stained for endothelial and immune markers (Lin+) (CD31, CD45) in PE, EPCAM in APC-Cy7 and were gated to eliminate debris (e) and to discard the doublets (f). The living cells were gated by Hoescht dye exclusion (g) and the non-epithelial Lin+ cells were discarded (h). EPCAM expression was studied in Lin-YFP+ (i). **j**, FACS quantification of the percentage of EPCAM^{+} cells in the different SCCs subtypes. (Medians and interquartile range (IQR) are shown, two-sided Mann-Whitney *U*-test, n represents the number of SCCs). **k**, Co-immunostaining of YFP and activated Caspase-3 or K14 in differentiated, mixed and mesenchymal tumours after 24h of cisplatin/5FU treatment (n = 10; scale bar, 20μm). Nuclei are stained with DAPI (blue). **l-o**, FACS plots showing the gating strategy to quantify the percentage of activated caspase-3 positive cells in YFP+ EPCAM^{+} and EPCAM^{-} cells after eliminating debris and discarding the doublets (Fig. 1a, d; 2c, e, i, j; 5d and Extended Data Fig. 2d, 7g, h, 8d, e, 10e, g). The same strategy (**l-n**) was used to quantify the percentage of cells expressing γH2AX and cell cycle analysis (Fig. 4c–e) in YFP+EPCAM^{+}, EPCAM^{-} and *Rhoj* KO EPCAM^{-} cells. **p**, FACS quantification of the percentage of activated caspase-3 positive cells in YFP+/EPCAM^{+} tumour cells, YFP+/EPCAM^{-} tumour cells and in the six different subpopulations of EPCAM^{-} tumour cells based on CD106/51/61 markers 24h after cisplatin/5FU administration to mice harbouring SCCs (Medians (IQR) are shown, Kruskal-Wallis test followed by two-sided Mann Whitney tests, n represents the number of SCCs).

Source data

Extended Data Fig. 2 RHOJ knockdown in EMT tumour cells decreases cell survival, growth and migration without affecting the expression of EMT or epithelial markers.

a, Relative mRNA expression of *Rhoj* and *Rhoq* by RNA seq in FACS isolated $EPCAM^+$ and - $EPCAM^-$ cells (Bar chart indicates mean; n = 2 SCCs). **b-c**, qRT-PCR of *Rhoj* and *Rhoq* expression validating the specific downregulation of *Rhoj* and *Rhoq* after shRNA knockdown in $EPCAM^-$ cells as compared to control shRNA $EPCAM^-$ cells (b) and after shRNA knockdown in $EPCAM^+$ cells as compared to control shRNA $EPCAM^+$ cells (c) (data are normalized using housekeeping gene *Tbp*, Bar chart indicates mean± SD, n represents the number of independent primary cell lines). **d**, FACS quantification of the percentage of activated caspase-3 positive cells in control $EPCAM^+$, *Rhoj* shRNA knockdown $EPCAM^+$ and *Rhoq* shRNA knockdown $EPCAM^+$ cells 24h after cisplatin/5FU administration (box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median, Kruskal-Wallis test followed by two-sided Mann Whitney tests, n represents the numbers of replicates from 2 independent cell lines). **e**, Relative cell number as compared to day 0 in control $EPCAM^-$ and *Rhoj* shRNA knockdown cells cultured in vitro and quantified by crystal violet assay (Medians and range are shown, two-sided Mann-Whitney *U*-test, n represents replicates of three biological replicates). **f-g**, Cell Migration. Representative images (f) and quantification (g) of cell migration following scratch-wound assay in $EPCAM^-$ control and *Rhoj* shRNA knockdown cells (scale bar, 500µm; data represent the percentage of wound closure, medians with range are shown, two-sided Mann-Whitney *U*-test, n = 6 biological replicates, 3–4 pictures per well). **h**, mRNA expression of selected epithelial and EMT markers quantified by qRT-PCR in $EPCAM^-$ control cells, *Rhoj* shRNA knockdown cells, $EPCAM^+$ control cells and *Rhoj* overexpressing cells (data are normalized using housekeeping gene TBP. Bars represent mean± SD, n represents the number of independent primary cell lines).

Source data

Extended Data Fig. 3 RHOJ overexpression in EPCAM$^{\pm}$ tumour cells in vitro and RHOJ deletion in tumour cells in vivo impair cell proliferation.

a-**b**, HA-tagged RHOJ expression measured by western blotting (a) and immunofluorescence (b) validating the overexpression (OE) of RHOJ in EPCAM^{+} cells as compared to control empty vector transduced EPCAM^{+} cells. Molecular weights (kDa) are indicated on the left side of the blots (n = 3 biological replicates; scale bar, 20µm). **c**, Relative cell number of YFP+ EPCAM^{+} control and RHOJ OE cells quantified by crystal violet assay (Medians and range are shown, two-sided Mann-Whitney *U*-test, n represents duplicates of two biological replicates). **d**, Scheme of the genetic strategy to induce *Kras*^{G12D} and YFP expression, *Trp53* and *Rhoj* deletion in the mouse model of skin SCCs. **e**, PCR analysis of *Cre* expression and *Rhoj* floxed allele recombination in *Lgr5*^{creER}*Kras*^{G12D}*Trp53*^{cKO}*Rosa-YFP* mice (n = 3), *Lgr5*^{creER}*Kras*^{G12D}*Trp53*^{cKO}*Rhoj*^{cKO}*Rosa-YFP* (*Rhoj* KO) mice (n = 3) and *Rhoj* KO tumours (n = 3) confirming the deletion of the floxed alleles in *Rhoj*-KO SCCs arising after tamoxifen administration. **f**, Hematoxylin and Eosin staining of WT and *Rhoj*-KO SCCs. (Scale bar, 50µm; n = 15 SCCs). **g**, Co-immunostaining of YFP (tumour cells), K14 and Ki67 (proliferating cells) in epithelial and mesenchymal part of *Rhoj* WT and *Rhoj*-KO SCCs (Scale bar, 100 µm). **h**, Quantification of the percentage of proliferating cells in epithelial (YFP+K14+) and mesenchymal (YFP+K14-) tumour cells from *Rhoj*-WT and *Rhoj*-KO SCCs (n = 7 independent tumour samples, Kruskal-Wallis test followed by two-sided Mann-Whitney *U*-tests, p-values adjusted by Bonferroni correction; box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median). **i**, Co-Immunostaining of active caspase-3 and YFP in primary tumours from WT and *Rhoj*-KO SCCs treated with cisplatin/5FU for 24h (n = 5; scale bar, 20µm).

Source data

Extended Data Fig. 4 RHOJ mediates resistance to chemotherapy associated with EMT in human cancer cells.

a, qRT-PCR of *Rhoj* and *Rhoq* expression validating the specific downregulation of *Rhoj* after shRNA knockdown (*Rhoj* sh) in human breast cancer cell line MDA-MB-231 as compared to control cells. (n = 3 independent experiments, data are normalized using housekeeping gene *Polr2a*; Control shRNA (Control) is an empty shRNA vector). **b**, Relative cell number as compared to day 0 in MDA-MB-231 control and RhoJ shRNA knockdown cultured cells in vitro and quantified by performing crystal violet assay (Medians and range are shown, two-sided Mann-Whitney *U*-test, n represents triplicates from two biological replicates). **c**-**d**, Representative images (c) and cell survival analysis (d) after 48h of treatment with 8, 16, 24 µM cisplatin; 30, 100, 300 nM paclitaxel; 250, 500, 700 nM doxorubicin of MDA-MB-231 control cells and *Rhoj* shRNA knockdown (scale bar, 500 µm; box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median, two-sided Mann-Whitney U-test, n represents replicates from two biological replicates).

Extended Data Fig. 5 Transcriptomic and proteomic characterization of EMT tumour cells after *Rhoj* deletion.

a, **b**, mRNA expression of genes upregulated (a) or downregulated (b) in EPCAM$^-$ Control shRNA compared to EPCAM$^-$ *Rhoj* sh KD measured by RNA seq. (Means are shown, n = 2 independent primary cultured cell lines). **c**,**d**, Gene Ontology analysis of genes that are upregulated (c) or downregulated (d) in EPCAM$^-$ control cells compared to *Rhoj* Sh KD EPCAM$^-$ cells (c), showing categories of genes that are significantly enriched. **e**-**f**, Gene Ontology analysis corresponding to the proteins significantly upregulated in EPCAM$^-$ WT and EPCAM$^-$ Control sh (e) and in EPCAM$^+$ and EPCAM$^-$ *Rhoj* sh (f). p value is calculated according to the Benjamini–Hochberg method for multiple hypothesis testing. **g**, Immunofluorescence of Phalloidin (red) in EPCAM$^+$, EPCAM$^-$ and EPCAM$^-$ *Rhoj* KO cells. Nuclei are counterstained with DAPI (blue) (n = 3 biological replicates; scale bar, 20µm). **h**-**k**, Western blot showing the expression of POLD (h), PCNA (i), phospho-RPA2 (S4/8), total RPA2 (j) and N-WASP (k) in EPCAM$^+$, EPCAM$^-$ and *Rhoj* KO cells. Tubulin or β-Actin loading controls (n = 2, molecular weights (kDa) are indicated to the

right side of the blots). **l**, mRNA expression of EMT transcription factors measured by RNA-sequencing in EPCAM^{-} control cells compared to EPCAM^{-} *Rhoj* sh KD cells (Means are shown, n represents the number of independent primary cell lines). **m**, Protein expression of selected epithelial and EMT markers in EPCAM^{+}, EPCAM^{-} WT, EPCAM^{-} Control sh, EPCAM^{-} *Rhoj* sh cells untreated and treated for 24h with cisplatin/5FU (n represents the number of technical replicates, Means + S.D. are shown).

Source data

Extended Data Fig. 6 RHOJ promotes DNA repair in EMT tumour cells independently of ATM and ATR.

a, Western blot analysis of phospho-ATM/ATR substrates (S*Q) in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells untreated, treated with cisplatin/5FU for 8 h and treated with cisplatin/5FU for 8 h followed by 16 h of ATM inhibitor (KU60019, 5µM) or ATR inhibitor (VE-821, 1µM) or a combination of ATM and ATR inhibitor showing the specific decrease in the level of phosphorylated ATM/ ATR substrates upon combination of their respective inhibitors with chemotherapy (n = 2 independent experiments, β-Actin loading control, molecular weights (kDa) are indicated to the right side of the blots). **b**, **c**, FACS quantification of the percentage of activated caspase-3 positive cells treated with cisplatin/5FU for 8h followed by treatment with ATR inhibitor (b) or ATM inhibitor (c) for 16h. Top panel shows drug treatment scheme. (two-sided Mann-Whitney *U*-test, n represents the number of replicates using 4 different primary cultured cell lines, box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median). **d**, **e**, Representative immunofluorescence (d), and FACS plots (e) of the cells expressing γH2AX in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells, 12h and 24h after cisplatin/5FU administration. (n = 3 independent primary cultured cell lines; scale bar, 10µm). **f**, Representative immunofluorescence and quantification of 53BP1, RPA2 and RAD51 focus formation (in red) in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells treated with cisplatin for 12h and 24h. (Scale bar, 10µm. Medians (IQR) are shown, Kruskal-Wallis test, p-values adjusted for multiple comparisons by Bonferroni correction, n

represents the number of analysed nuclei from two independent primary cultured cell lines and experiments).

Extended Data Fig. 7 RHOJ promotes tolerance to replicative stress in EMT tumour cells following chemotherapy.

a, Immunofluorescence of EdU incorporation (green) in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells treated with cisplatin/5FU for 12 and 24h. EdU pulse is given 30 min before sampling. Nuclei are counterstained with DAPI (blue) (Scale bar, 25µm). **b**, EdU intensity measured with Cell profiler software in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells treated with cisplatin/5FU for 12 and 24h. (n represents the number of nuclei analysed from 2 independent experiments ; a.u., arbitrary unit, Medians are shown, Kruskal-Wallis test) **c**-**f**, Western blot of p-CDK1 at Y15 (c), at Thr161 and total CDK1 (d), p-CDK2 at Thr160 and total CDK2 (e), p-CDK4 at Thr172, total CDK4 and CDK6 (f) and corresponding β-Actin in EPCAM^{+}, EPCAM^{-} and *Rhoj* KO cells exposed to cisplatin/5FU treatment showing the differential activation of the CDKs in response to chemotherapy in the different tumour cell types. Activation of CDK1 was observed in EPCAM^{-} cells as shown by transient level of inhibitory phosphorylation and earlier appearance of activating phosphorylation of CDK1 in response to chemotherapy as compared to EPCAM^{+} and *Rhoj* KO Epcam- cells. Same levels of activating CDK2 phosphorylation was found in all cell types and sustained activating CDK4 phosphorylation was observed in EPCAM^{-} cells 12h after chemotherapy compared to EPCAM^{+} and *Rhoj* KO EPCAM^{-} cells suggesting that EPCAM^{-} cells are allowed to progress in the cell cycle following chemotherapy. Molecular weights (kDa) are indicated on the right side of the blots (n = 2). **g**, FACS quantification of the percentage of activated caspase-3 positive cells in YFP+ EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO in response to 24h aphidicolin (50µM) and cisplatin/5FU. **h**, FACS quantification of the percentage of activated caspase-3 positive cells in response to 24h combination of cisplatin/5FU with MRE11 inhibitor (Mirin 50µM). Lower panel shows drug treatment scheme. (n represents the number of independent primary cultured cell lines). In **g** and **h**, Kruskal-Wallis test followed by Mann-Whitney *U*-tests. p-values adjusted for multiple comparisons by Bonferroni correction, box

boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median.

Extended Data Fig. 8 Replicative stress in EMT tumour cells is associated with the activation of dormant origins in a RHOJ-dependent manner that prevents micronuclei formation.

a-**b**, Representative images (a) and quantification (b) of micronuclei in $EPCAM^+$, $EPCAM^-$ and $EPCAM^-$ *Rhoj* KO cells untreated and treated with cisplatin/5FU for 12 and 24h. Nuclei are stained with DAPI. Yellow arrowheads indicate micronuclei (a) (Scale bar, 25µm, n represents the number of nuclei pooled from five independent experiments, Kruskal-Wallis test followed by two-sided Mann-Whitney paired comparisons tests. p-values adjusted for multiple comparisons by Bonferroni correction, box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median). **c**-**d**, Representative images (c) and quantification (d) of fork asymmetry to measure fork stalling in$EPCAM^+$, $EPCAM^-$ and $EPCAM^-$ *Rhoj* KO cells untreated and treated with cisplatin/5FU for 12 h. The degree of symmetry around the replication origin was calculated as long/short fork ratio (100 individual forks were measured for each cell line in three replicates, medians with interquartile range are shown, ANOVA with condition, experience and their interaction, p-values after two-way post-hoc Sidak tests). **e**, Western blot analysis of the indicated replication factors and loading control in the whole cell extract and loaded on chromatin in untreated conditions and 12h after cisplatin/5FU administration (n = 2, molecular weights (kDa) are indicated on the right side of the blots).

Extended Data Fig. 9 Patterns of nuclear actin filaments in response to chemotherapy, RHOJ deletion and actin polymerization inhibitor.

a, Representative images of $EPCAM^-$ cells transfected with HA-tagged RHOJ untreated and 12h after cisplatin/5FU treatment. Maximum intensity projections are shown (n = 3 biological replicates; anti-HA in red; nuclei are counterstained with DAPI in green; scale bar, 20µm) (top). Corresponding

colocalization map of the overlap between HA-tagged RHOJ and DAPI signals showing peri- and intra- nuclear localization of RHOJ. Global intersection coefficient is represented (bottom). **b**, Representative images of the different nuclear actin patterns found in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells expressing nuclear actin chromobody-GFP treated with cisplatin/5FU and a combination of cisplatin/5FU + Latrunculin B (LatB, 400nM) or SMIFH2 (50 µM) for 12h. Nuclei are counterstained with DAPI (blue) (n = 3; scale bar, 10µm). **c**, Quantification of nuclear actin patterns in EPCAM^{+}, EPCAM^{-} and EPCAM^{-} *Rhoj* KO cells treated with cisplatin/5FU and a combination of cisplatin/5FU + Latrunculin B or SMIFH2 for 12h (Means are shown, n represents the number of independent primary cultured cell lines). Several classes of nuclear actin filaments are observed (b). Pattern 1, thin, short and branched filaments, is found in all cell types as shown in panel highlighted by a purple border ; pattern 2, elongated filaments, is mainly induced in EPCAM^{-} cells upon chemotherapy treatment as shown in panel highlighted by an orange border; pattern 3, hairy filaments which are short filaments with a dense and multipolar organization, is detected in EPCAM^{-} *Rhoj* KO cells, as shown in panel highlighted by a blue border; pattern 4, thick and twisted actin filaments as shown in panel highlighted by a yellow border and pattern 5, severe disruption of actin filaments as shown in panel highlighted by a green border, are found when chemotherapy was combined with F-actin inhibitors.

Extended Data Fig. 10 Nuclear actin filament formation occurs in replicative cells and participates in DNA repair and cell survival in response to chemotherapy.

a-**b**, Representative immunofluorescence of phospho-histone H3 (p-H3) (red) and EdU (white) in EPCAM^{-} (a) and EPCAM^{-} *Rhoj* KO (b) cells expressing nuclear actin chromobody-GFP treated with cisplatin/5FU or a combination of cisplatin/5FU and Latrunculin B for 12h compared to untreated conditions. Nuclei are counterstained with DAPI (blue). Arrowhead indicates a nucleus presenting co-localization of nuclear actin filament and EdU (Scale bar, 10µm). **c**, Quantification of EdU positive/ p-H3 negative cells in EPCAM^{-} and EPCAM^{-} *Rhoj* KO tumour cells presenting nuclear actin filaments of (a) and (b) (>100 cells counted per

condition; n represents the number of replicates from 2 independent cell lines, medians are shown, two-sided Mann-Whitney U-tests). **d**, Quantification of the proportion of each nuclear actin patterns in EPCAM$^-$ and EPCAM$^-$ *Rhoj* KO cells presenting nuclear actin filaments of (a) and (b) (Means are shown). **e**, FACS quantification of the percentage of activated caspase-3 positive cells in EPCAM$^+$, EPCAM$^-$ and EPCAM$^-$ *Rhoj* KO in response to Latrunculin B or SMIFH2 for 24h. (Medians are shown, Kruskal-Wallis test, n represents the number of independent primary cultured cell lines). **f**, Quantification of RPA2 foci in EPCAM$^-$ cells treated with a combination of cisplatin/5FU with Latrunculin B or SMIFH2 inhibitor for 12h. (Medians are shown, Kruskal-Wallis test, p-values adjusted for multiple comparisons by Bonferroni correction, n represents the number of analysed nuclei pooled from two independent primary cultured cell lines). **g**, FACS quantification of the percentage of activated caspase-3 positive cells in EPCAM$^-$ tumour cells in response to cisplatin/5FU combined with ARP2/3 inhibitor (CK666, 50 µM) or N-WASP inhibitor (Wiskostatin, 5 µM) for 24h (box boundaries represent 25th and 75th percentiles; whiskers, minimum and maximum; centre line, median; n represents the number of replicates from 3 independent cell lines, Kruskal-Wallis test followed by two- sided Mann Whitney tests). For gel source data, see Supplementary Figure 1 and 2.

Supplementary information

Supplementary Information

Supplementary Figs. 1 and 2. Supplementary Fig. 1: unprocessed original scans of Western blot analysis. Each panel indicates the figure to which the full membranes belong. Molecular mass size standards are indicated on each membrane. Supplementary Fig. 2: unprocessed original scans of western blot analyses. Each panel indicates the figure to which the full membranes belong (a–e). Molecular mass size standards are indicated on each membrane. Unprocessed original gel picture of PCR genotype analysis (f).

Reporting Summary

Peer Review File

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 5

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 5

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Debaugnies, M., Rodríguez-Acebes, S., Blondeau, J. *et al.* RHOJ controls EMT-associated resistance to chemotherapy. *Nature* **616**, 168–175 (2023). https://doi.org/10.1038/s41586-023-05838-7

- Received: 19 September 2021
- Accepted: 14 February 2023
- Published: 22 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05838-7

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Cancer cells remodel nuclear actin filaments to resist chemotherapy**

- Stephanie Panier

Nature (2023)

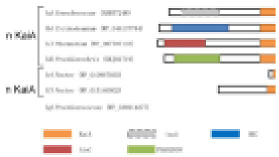
[Cancer cells remodel nuclear actin filaments to resist chemotherapy]

- Stephanie Panier

Nature News & Views 22 Mar 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05838-7

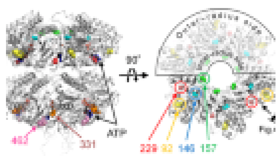
| Section menu | Main menu |



Evolution of kaiA, a key circadian gene of cyanobacteria

11 May 2021

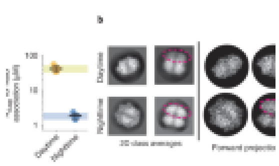
Volodymyr Dvornyk & Qiming Mei



Conformational rearrangements of the C1 ring in KaiC measure the timing of assembly with KaiB

11 June 2018

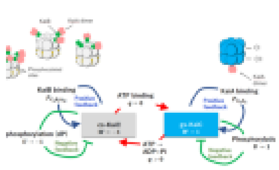
Atsushi Mukaiyama, Yoshihiko Furuike, … Shuji Akiyama



Coupling of distant ATPase domains in the circadian clock protein KaiC

21 July 2022

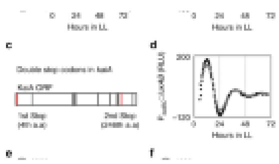
Jeffrey A. Swan, Colby R. Sandate, … Carrie L. Partch



Mechanism of autonomous synchronization of the circadian KaiABC rhythm

25 February 2021

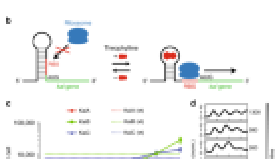
Masaki Sasai



Damped circadian oscillation in the absence of KaiA in Synechococcus

07 May 2020

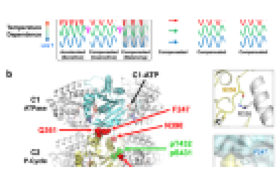
Naohiro Kawamoto, Hiroshi Ito, … Hideo Iwasaki



High protein copy number is required to suppress stochasticity in the cyanobacterial circadian clock

01 August 2018

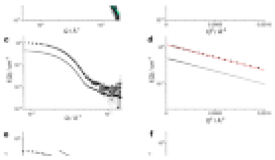
Justin Chew, Eugene Leypunskiy, … Michael J. Rust



Cross-scale analysis of temperature compensation in the cyanobacterial circadian clock system

04 April 2022

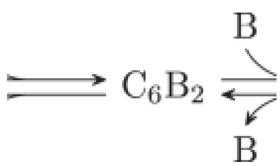
Yoshihiko Furuike, Dongyan Ouyang, … Shuji Akiyama



[Overall structure of fully assembled cyanobacterial KaiABC circadian clock complex by an integrated experimental-computational approach]

10 March 2022

Yasuhiro Yunoki, Atsushi Matsumoto, … Masaaki Sugiyama



[An alternative interpretation of the slow KaiB-KaiC binding of the cyanobacterial clock proteins]

26 June 2020

Shin-ichi Koda & Shinji Saito

- Article
- Open Access
- Published: 22 March 2023

From primordial clocks to circadian oscillators

- Warintra Pitsawong,
- Ricardo A. P. Pádua,
- Timothy Grant,
- Marc Hoemberger,
- Renee Otten,
- Niels Bradshaw,
- Nikolaus Grigorieff &
- …
- Dorothee Kern

Show authors

Nature **volume 616**, pages 183–189 (2023)

- 11k Accesses
- 77 Altmetric
- Metrics details

Abstract

Circadian rhythms play an essential part in many biological processes, and only three prokaryotic proteins are required to constitute a true post-translational circadian oscillator[1]. The evolutionary history of the three Kai proteins indicates that KaiC is the oldest member and a central component of the clock[2]. Subsequent additions of KaiB and KaiA regulate the phosphorylation state of KaiC for time synchronization. The canonical KaiABC system in cyanobacteria is well understood[3,4,5,6], but little is known about more ancient systems that only possess KaiBC. However, there are reports that they might exhibit a basic, hourglass-like timekeeping mechanism[7,8,9]. Here we investigate the primordial circadian clock in *Rhodobacter sphaeroides*, which contains only KaiBC, to elucidate its inner workings despite missing KaiA. Using a combination of X-ray crystallography and cryogenic electron microscopy, we find a new dodecameric fold for KaiC, in which two hexamers are held together by a coiled-coil bundle of 12 helices. This interaction is formed by the carboxy-

terminal extension of KaiC and serves as an ancient regulatory moiety that is later superseded by KaiA. A coiled-coil register shift between daytime and night-time conformations is connected to phosphorylation sites through a long-range allosteric network that spans over 140 Å. Our kinetic data identify the difference in the ATP-to-ADP ratio between day and night as the environmental cue that drives the clock. They also unravel mechanistic details that shed light on the evolution of self-sustained oscillators.

Main

Circadian clocks are self-sustained biological oscillators that are ubiquitously found in prokaryotic and eukaryotic organisms. In eukaryotes, these systems are complex and highly sophisticated, whereas in prokaryotes, the core mechanism is regulated by a post-translational oscillator that can be reconstituted in vitro with ATP and three proteins (encoded by *kaiA*, *kaiB* and *kaiC*)[1]. Seminal work on the KaiABC system has resulted in a comprehensive understanding of its circadian clock. KaiC is the central component that autophosphorylates by binding to KaiA and autodephosphorylates following association with KaiB[3,4,5,6]. The interplay among these three proteins has been shown in vitro to constitute a true circadian oscillator characterized by persistence, resetting and temperature compensation. Consequently, the KaiABC system is considered an elegant and the simplest implementation of a circadian rhythm. The evolutionary history of *kai* genes established *kaiC* as the oldest member dating back around 3.5 billion years ago. Subsequent additions of *kaiB* and most recently *kaiA* formed the extant *kaiBC* and *kaiABC* clusters, respectively[2,10]. Notably, some studies of more primitive organisms that lack *kaiA* hinted that the *kaiBC*-based systems might already provide a basic, hourglass-like timekeeping mechanism[7,8,9]. Contrary to the self-sustained oscillators found in cyanobacteria, such a timer requires an environmental cue to drive the clock and for the daily flip of the hourglass. The central role of circadian rhythms in many biological processes, controlled by the day and night cycle on Earth, makes their evolution a fascinating topic.

Here we investigate such a primitive circadian clock through biochemical and structural studies of the KaiBC system of the purple, nonsulfur photosynthetic proteobacterium *R. sphaeroides* KD131 (hereafter, its

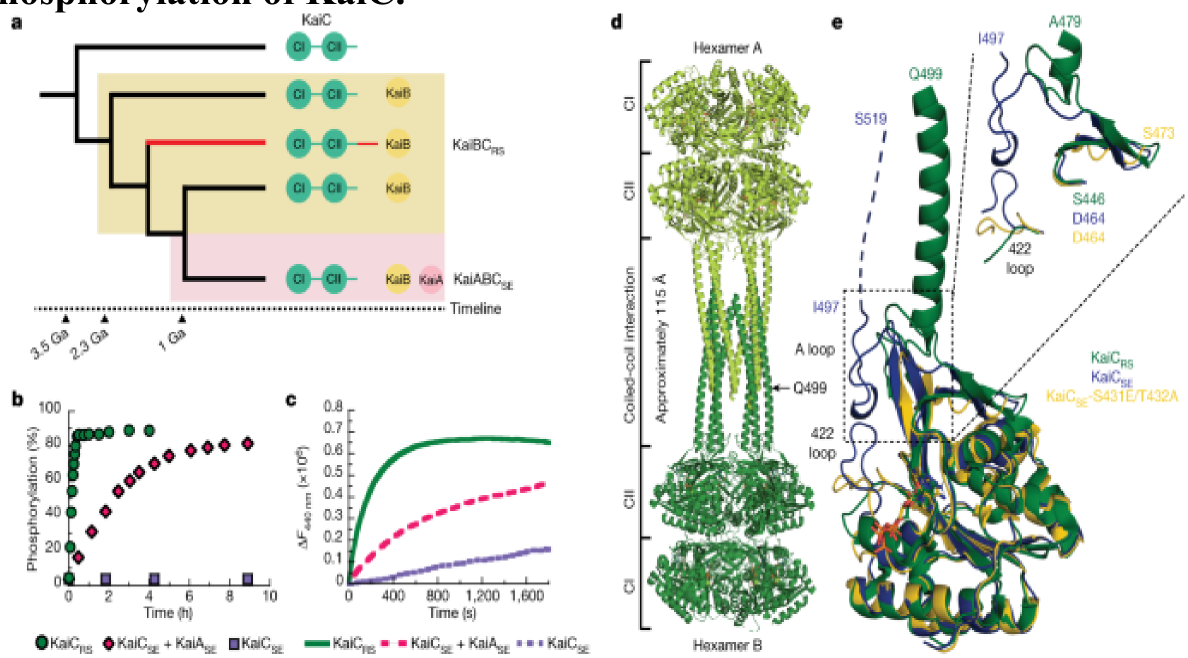
components are referred to as $KaiB_{RS}$ and $KaiC_{RS}$). The organism shows sustained rhythms of gene expression in vivo, but whether *kaiBC* is responsible for this observation remains inconclusive in the absence of a *kaiC* knockout[11]. A previous study of the closely related bacterium *Rhodopseudomonas palustris* that used a knockout strain demonstrated causality between the proto-circadian rhythm of nitrogen fixation and expression of the *kaiC* gene[9]. Here through in vitro experiments, we discover that $KaiBC_{RS}$ is a primordial circadian clock with a mechanism that is different from the widely studied circadian oscillator in *Synechococcus elongatus* PCC 7942 (hereafter, its components are referred to as $KaiA_{SE}$, $KaiB_{SE}$ and $KaiC_{SE}$)[3,4,5,6]. We identify an environmental cue that regulates the phosphorylation state and consequently produces a 24 h clock in vivo as the switch in the ATP-to-ADP ratio between day and night. Our results from kinetic studies combined with X-ray and cryogenic electron microscopy (cryo-EM) structures of the relevant states unravel a long-range allosteric pathway that is crucial for the function of the hourglass and sheds light on the evolution of self-sustained oscillators. Notably, we find a new protein fold for $KaiC_{RS}$ and uncover a register shift in the coiled-coil domain that spans around 115 Å as the key regulator in this system, which shows structural similarities to dynein signalling[12].

The C-terminal tail is a primitive regulatory moiety

To gain insight into the evolution of the *kaiBC* cluster, we constructed a phylogenetic tree of *kaiC* after the emergence of *kaiB* (Fig. 1a, Extended Data Fig. 1a and Supplementary Datasets 1 and 2). The first question we sought to answer is how $KaiC_{RS}$ and other members in the clade can autophosphorylate despite having no KaiA. KaiA is known to be crucial for this function in the canonical KaiABC system at its optimum temperature. We observed a large clade that exhibits a C-terminal tail about 50 amino acids longer compared with *kaiC* in other clades (Extended Data Fig. 1b). This C-terminal extension near the A loop is predominantly found in the *kaiC2* subgroup, which was previously annotated as having two serine phosphorylation sites instead of the Thr–Ser pair found in the *kaiC1* and

kaiC3 subgroups[13,14,15] (Extended Data Fig. 1b). In *S. elongatus*, the binding of $KaiA_{SE}$ to the A loop of $KaiC_{SE}$ tethers them in an exposed conformation[16] that activates both autophosphorylation and nucleotide exchange[17]. Given the proximity of the extended C-terminal tail to the A loop, we conjectured that it could serve as the 'primitive' regulatory moiety that was made redundant with the appearance of KaiA.

Fig. 1: The extended C-terminal tail of $KaiC_{RS}$ forms a coiled-coil interaction with an exposed A loop for KaiA-independent phosphorylation of KaiC.



a, Schematic of the phylogenetic tree of *kaiC* showing the appearance of *kaiB* and *kaiA* during evolution. The *kaiC* clade with an approximately 50-amino-acid C-terminal extension is labelled in red, and a timeline was predicted as previously reported[2]. Ga, billion years ago. **b**, Phosphorylation rate over time of $KaiC_{RS}$ (6.5 ± 1.0 h^{-1}) and $KaiC_{SE}$ in the presence (0.40 ± 0.02 h^{-1}) or absence of $KaiA_{SE}$ at 30 °C. The s.d. in reported parameters were obtained from the fitting. **c**, Nucleotide exchange between ATP and mant-ATP in $KaiC_{RS}$ alone (18.0 ± 1.5 h^{-1}) compared with $KaiC_{SE}$ in the presence (4.7 ± 0.3 h^{-1}) and absence of $KaiA_{SE}$ (0.08 ± 0.04 h^{-1}) measured at 30 °C. Representative traces are shown, and the fitted parameters (mean ± s.d.) were obtained from three replicate measurements. **d**, X-ray structure of

dodecameric $KaiC_{RS}$ (PDB: 8DBA) coloured by hexamer A (light green) and hexamer B (dark green). The CI, CII and coiled-coil domains are indicated, and the A loop is labelled in **e**. **e**, Superposition based on an alignment of the CII domain of $KaiC_{RS}$ (green; PDB: 8DBA, chain B), $KaiC_{SE}$ (purple; PDB: 1TF7, chain B)[36] and $KaiC_{SE}$-S431E/T432A (yellow; PDB: 7S65, chain A)[19] shows that $KaiC_{RS}$ has an extended A loop orientation that no longer forms the inhibitory interaction with the 422 loop ($KaiC_{SE}$ numbering). The conformation of the 422 loop in $KaiC_{RS}$ resembles the one seen in the cryo-EM structure of the phosphomimetic $KaiC_{SE}$-S431E/T432A (yellow; PDB: 7S65)[19]. No electron density is observed for the C-terminal part of wild-type $KaiC_{SE}$ and the S431E/T432A mutant owing to flexibility, and the missing 22 residues for wild-type $KaiC_{SE}$ (46 for S431E/T432A) are represented by a dashed line (not shown for the mutant).

To test our hypothesis, we first measured the autophosphorylation and nucleotide exchange rates in $KaiC_{RS}$, which both depend on the presence of KaiA in the $KaiABC_{SE}$ system. We observed an autophosphorylation rate for $KaiC_{RS}$ that was about 16-fold higher than for $KaiC_{SE}$ activated by $KaiA_{SE}$ ($6.5 \pm 1.0\ h^{-1}$ compared with $0.40 \pm 0.02\ h^{-1}$, respectively; Fig. 1b and Extended Data Fig. 2a–e). Similarly, the nucleotide exchange rate was faster in $KaiC_{RS}$ compared with $KaiC_{SE}$, even in the presence of $KaiA_{SE}$ ($18.0 \pm 1.5\ h^{-1}$ compared with $4.7 \pm 0.3\ h^{-1}$, respectively; Fig. 1c and Extended Data Fig. 2f). Our data show that $KaiC_{RS}$ can perform both autophosphorylation and nucleotide exchange on its own and does so faster than its more recently evolved counterparts.

A coiled-coil interaction assembles a $KaiC_{RS}$ dodecamer

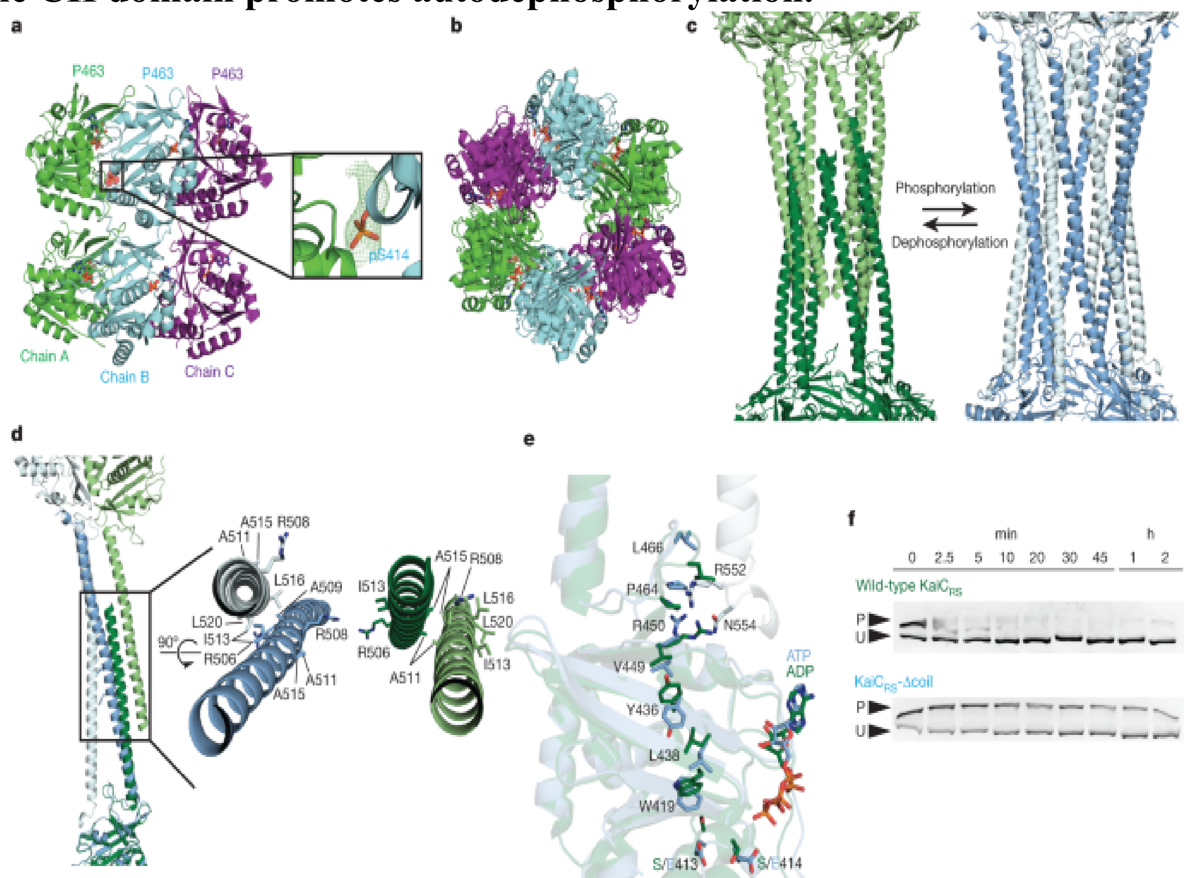
To mechanistically assess how KaiC in *kaiA*-null systems accomplishes autophosphorylation, we turned to structural biology. The crystal structure of $KaiC_{RS}$, unlike KaiC from cyanobacteria, revealed a homododecamer that

consisted of two homohexameric domains joined by a 12-helical coiled-coil domain that is formed by the extended C-terminal tail (Protein Data Bank (PDB) identifier: 8DBA; Fig. 1d and Extended Data Table 1). A closer inspection of the CII domains in $KaiC_{RS}$ and $KaiC_{SE/TE}$ (*Thermosynechococcus elongatus* BP-1 referred to as $KaiC_{TE}$) showed an obvious difference in A loop orientations: an extended conformation in $KaiC_{RS}$ compared with a buried orientation in $KaiC_{SE/TE}$ (Fig. 1e). The existence of such an extended conformation following binding of KaiA has been previously proposed[18]. This hypothesis was based on the perceived hyperphosphorylation and hypophosphorylation that occurred after removing the A loop or disrupting KaiA binding, respectively[18]. A recently solved cryo-EM structure of the night-time phosphomimetic $KaiC_{SE}$-S431E/T432A in its compressed state directly showed a disordered A loop that no longer interacts with the 422 loop[19], similar to the extended A loop conformation we observed in $KaiC_{RS}$ (Fig. 1e). The loss of interaction between the A loop and the 422 loop (just 10 residues apart from the phosphorylation sites) results in closer proximity between the hydroxyl group of Ser431–Thr432 and the γ-phosphate of ATP, thereby, facilitating the phosphoryl transfer step[20]. Furthermore, the sequence similarity between $KaiC_{RS}$ and $KaiC_{SE}$ is less than 30% for the A loop and residues considered important for stabilization of this loop in its buried orientation (that is, the 422 loop and residues 438–444) (Fig. 1e). Together, our structural and kinetic data support the idea that an exposed A loop is key for the KaiA-independent enhancement of nucleotide exchange and hence autophosphorylation in $KaiC_{RS}$ and perhaps other KaiBC-based systems.

We then questioned whether the purpose of the coiled-coil domain is to ‘pull up’ the A loop or to actively participate in nucleotide exchange and autophosphorylation of KaiC. To further understand its role, we generated a truncation at residue Glu490 based on the phylogenetic tree and crystallographic information ($KaiC_{RS}$-Δcoil) (Extended Data Fig. 1b) to disrupt the coiled-coil interaction between the two hexamers. The crystal structure of $KaiC_{RS}$-Δcoil (PDB: 8DB3; Fig. 2a,b and Extended Data Table 1), its size-exclusion chromatogram and analytical ultracentrifugation profile (Extended Data Fig. 3a–c) showed a hexameric structure with no coiled-coil

interaction. Nucleotide exchange rates in the CII domain for $KaiC_{RS}$-Δcoil and the wild-type protein were comparable ($19.1 \pm 0.8\ h^{-1}$ and $18.0 \pm 1.5\ h^{-1}$, respectively; Extended Data Fig. 3d). The phosphorylation rates were also similar ($5.5 \pm 0.4\ h^{-1}$ and $7.4 \pm 0.3\ h^{-1}$ for $KaiC_{RS}$-Δcoil and wild type, respectively; Extended Data Fig. 3e,f). These results indicate that the extended A loop and not the coiled-coil interaction plays a pivotal part in nucleotide exchange and autophosphorylation in $KaiC_{RS}$. The results also provide a potential mechanism of autophosphorylation in other KaiBC-based systems that lack a coiled-coil bundle. Notably, the coiled-coil bundle provides additional hexameric stability. In detail, the $KaiC_{RS}$ dodecamer is stable for extended periods of time in the presence of only ADP (Extended Data Fig. 3g,h), whereas for $KaiC_{SE}$, oligomers are not observed under these conditions[21].

Fig. 2: A coiled-coil partner switch coupled to an allosteric network in the CII domain promotes autodephosphorylation.



a, X-ray structure of $KaiC_{RS}$-Δcoil was solved in the $C222_1$ space group and contained three monomers in the asymmetric unit, with ADP present in all active sites. The obtained electron density map allowed for model building up to Pro463, which indicated that the truncation at position 490 results in enhanced flexibility beyond Pro463. Phosphorylation of Ser414 (pS414) was observed in chain B (cyan) as shown by the electron density $mF_o–DF_c$ polder map (green mesh, 3σ contour level). **b**, Assembly analysis using the PISA software[37] revealed a hexamer as the most probable quaternary structure (top view). **c**, Structural comparison of the coiled-coil domain for unphosphorylated $KaiC_{RS}$ (dark and light green; X-ray structure) and the $KaiC_{RS}$-S413E/S414E phosphomimetic mutant (dark and light blue; cryo-EM structure). **d**, Overlay of interacting dimers of the structures in **c** using the CII domain of chain A as a reference (dark shades; bottom). Unphosphorylated $KaiC_{RS}$ (dark green) interacts with the opposite partner on the right (light green), whereas $KaiC_{RS}$-S413E/S414E (dark blue) interacts with the partner on the left (light blue). The hydrophobic packing in the coiled-coil domain is mediated by only the Cβ atoms of alanine and arginine residues in unphosphorylated $KaiC_{RS}$, but involves the entire side chain of leucine and isoleucine residues in the phosphomimetic structure. **e**, Allosteric network in the phosphomimetic state (blue) from the coil (light blue) propagating through the $KaiC_{RS}$ CII domain to the active site (dark blue) compared with the unphosphorylated state (dark green) (Supplementary Video 1). **f**, Autodephosphorylation of $KaiC_{RS}$ and $KaiC_{RS}$-Δcoil over time in the presence of 4 mM ADP at 30 °C. The phosphorylated (P) and unphosphorylated (U) proteins were separated by Zn^{2+} Phos-tag SDS–PAGE (for gel source data, see Supplementary Fig. 1).

A long-range allosteric network in $KaiC_{RS}$

The change in phosphorylation state of KaiC has been well established to be the central feature for the circadian rhythm[22,23]. Notably, when comparing the unphosphorylated form of full-length $KaiC_{RS}$ (PDB: 8DBA) and its phosphomimetic mutant (S413E/S414E; PDB: 8FWI) (Extended Data Fig. 4 and Extended Data Table 2), we observed two distinct coiled-coil interactions. Following phosphorylation, the coiled-coil pairs swap partners

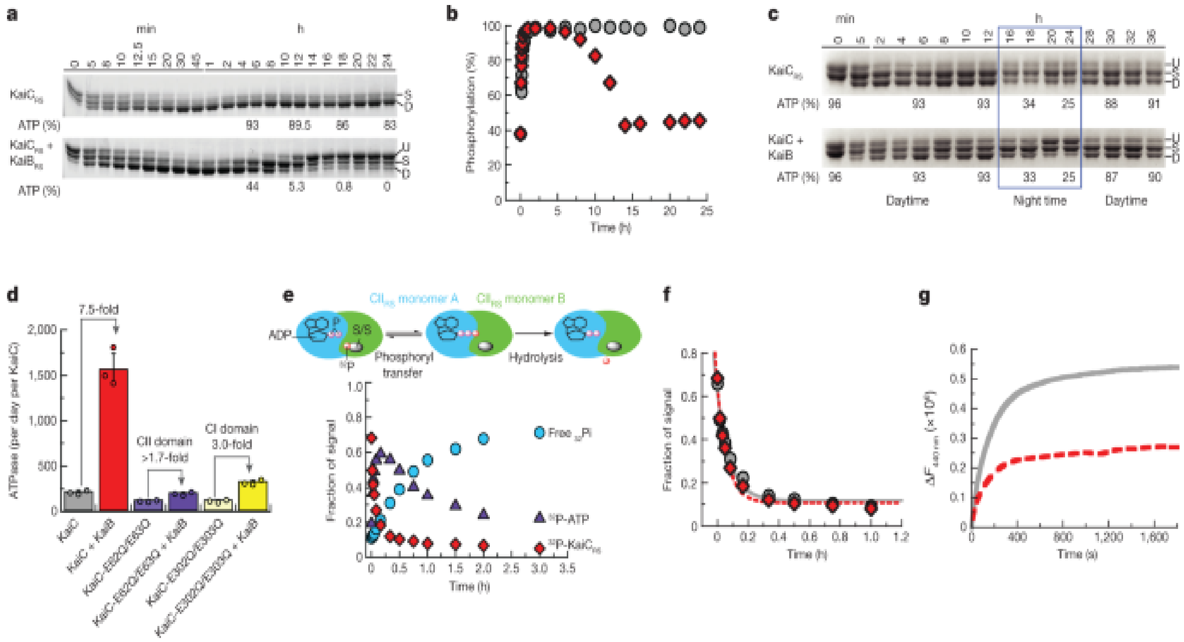
by interacting with the other neighbouring chain from the opposite hexamer, which resulted in a register shift that propagated around 115 Å along the entire coiled-coil (Fig. 2c and Extended Data Fig. 5). In the phosphomimetic state, the register comprised bulkier hydrophobic residues that resulted in a more stable interaction than for the dephosphorylated form (Fig. 2d and Extended Data Fig. 3g). Furthermore, the C-terminal residues of $KaiC_{RS}$-S413E/S414E interacted with the CII domain of the opposite hexamer, whereas the lack of electron density for the last 30 residues in the wild-type structure indicates more flexibility in the dephosphorylated state. We discovered that these conformational changes in the coiled-coil domain seemed to be coupled through a long-range allosteric network to the phosphorylation sites. The rotameric states of residues Ser413, Ser414, Trp419, Val421, Tyr436, Leu438, Val449 and Arg450 moved concertedly and pointed towards the nucleotide-binding site when the protein was phosphorylated or pointed away in the absence of a phosphate group (Fig. 2e, Extended Data Fig. 5d and Supplementary Video 1). We propose that the proximity of the nucleotide to the phosphorylated residue facilitated more efficient phosphoryl transfer. We therefore experimentally determined the impact of the coiled-coil domain on the autodephosphorylation rate of $KaiC_{RS}$. The wild-type protein dephosphorylated comparatively quickly (observed rate constant = $11.5 \pm 0.8\ h^{-1}$) in the presence of only ADP. By contrast, little dephosphorylation was observed for $KaiC_{RS}$-Δcoil (Fig. 2f and Extended Data Fig. 3i), for which allosteric propagation was disrupted (Extended Data Fig. 5d). Consistent with this accelerated dephosphorylation rate mediated by the coiled-coil domain, our crystallographic data showed a phosphate group on Ser414 for $KaiC_{RS}$-Δcoil but not for the wild-type protein (Fig. 2a and Extended Data Fig. 5d).

The ATP-to-ADP ratio resets the clock

It was notable that $KaiC_{RS}$ can autodephosphorylate on its own despite being constitutively active for phosphorylation owing to its extended A loop conformation. In the canonical *kaiABC* system, the interaction between KaiB and KaiC is required to provide a new binding interface that sequesters KaiA from its activating binding site, thereby promoting autodephosphorylation at the optimum temperature of the organism[24,25,26].

We therefore sought to discover whether the $KaiC_{RS}$ system can oscillate and whether there is a regulatory role for $KaiB_{RS}$ in this process. Comparing the in vitro phosphorylation states of $KaiC_{RS}$ in the absence and presence of $KaiB_{RS}$ showed an initial, rapid phosphorylation followed by an oscillatory-like pattern in the presence of $KaiB_{RS}$ (hereafter referred to as $KaiBC_{RS}$), whereas $KaiC_{RS}$ alone remained phosphorylated (Fig. 3a,b). Notably, the ATP consumption during the reaction with $KaiB_{RS}$ was significantly higher than without (Fig. 3a). As noted above, $KaiC_{RS}$ will also dephosphorylate completely in the presence of only ADP (Fig. 2f). These results suggest that the phosphorylation state of $KaiC_{RS}$ and the observed oscillatory half-cycle (Fig. 3a,b) is probably related to a change in the ATP-to-ADP ratio. We conjectured that this could constitute the environmental cue to reset the timer. To test our hypothesis, an ATP-recycling system was added after complete dephosphorylation of $KaiBC_{RS}$. As predicted, $KaiC_{RS}$ was able to restart the cycle and phosphorylate again (Extended Data Fig. 6a). We note that in vivo, the ATP-to-ADP ratio will not vary as substantially as in this in vitro experiment, as nucleotide homeostasis is tightly regulated. To mimic the day and night period for *R. sphaeroides*, we repeated the experiments while keeping the ATP-to-ADP ratio constant (mostly ATP at daytime owing to photosynthesis compared with 25:75% ATP-to-ADP during night time)[27]. In the presence of high ATP (that is, mimicking daytime), $KaiC_{RS}$ remained single or double phosphorylated (Fig. 3c and Extended Data Fig. 6b) irrespective of $KaiB_{RS}$. By contrast, a constant 25:75% ATP-to-ADP ratio (that is, mimicking night time) resulted in a much higher fraction of dephosphorylated $KaiC_{RS}$ in the presence of $KaiB_{RS}$ (Fig. 3c). Moreover, when the ATP-to-ADP ratio was flipped to mimic daytime, $KaiC_{RS}$ was able to phosphorylate again (Fig. 3c, around the 28 h mark). Our data support the notion that the phosphorylation behaviour of $KaiBC_{RS}$ strongly depends on the ATP-to-ADP ratio and demonstrate that the physical binding of $KaiB_{RS}$ results in a higher level of $KaiC_{RS}$ dephosphorylation at night time.

Fig. 3: The regulatory role of $KaiB_{RS}$ in the phosphorylation–dephosphorylation cycle of $KaiC_{RS}$.



a, SDS–PAGE gel of 3.5 μM $KaiC_{RS}$ and 4 mM ATP in the absence (top) and presence (bottom) of 3.5 μM $KaiB_{RS}$ at 35 °C, with the percentage of ATP indicated at specific time points. **b**, Phosphorylation (single and double) of $KaiC_{RS}$ during the reaction in the absence (grey circles) or presence (red diamonds) of $KaiB_{RS.}$ **c**, Phosphorylation–dephosphorylation cycle of 3.5 μM phosphorylated $KaiC_{RS}$ in the absence and presence of 3.5 μM $KaiB_{RS}$ in a constant ATP-to-ADP ratio of high ATP (4 mM) to mimic daytime and about 25% ATP to mimic the night time (exact percentage of ATP indicated at specific time points) at 30 °C. U, S and D in **a** and **c** represent the unphosphorylated, single phosphorylated (at Ser413 or Ser414) and double phosphorylated state of $KaiC_{RS}$, respectively (for gel source data, see Supplementary Fig. 1). **d**, ATPase activity of wild-type $KaiC_{RS}$ in the absence and presence of $KaiB_{RS}$, $KaiC_{RS}$-E62Q/E63Q in the absence and presence of $KaiB_{RS}$, and $KaiC_{RS}$-E302Q/E303Q in the absence and presence of $KaiB_{RS}$ at 30 °C. Bar graphs show mean ± s.d. from three replicates. **e**, Time-dependent autodephosphorylation of ^{32}P-labelled $KaiC_{RS}$ bound with ADP in the presence of 20 μM $KaiB_{RS}$ and 4 mM ADP at 30 °C showing phosphorylated ^{32}P-$KaiC_{RS}$, ^{32}P-ATP and free ^{32}Pi. The reaction products were separated by thin layer chromatography. **f**, The decay of phosphorylated ^{32}P-$KaiC_{RS}$ bound with 4 mM ADP in the absence (grey

circles) and presence (red diamonds) of $KaiB_{RS}$ at 30 °C is obtained from autoradiography quantification (Extended Data Fig. 7). **g**, The nucleotide exchange of 3.5 μM $KaiC_{RS}$ (grey trace) and 3.5 μM $KaiC_{RS}$ in complex with 30 μM $KaiB_{RS}$ (red dotted trace) in the presence of ATP with mant-ATP. Representative traces are shown, and the fitted parameters (mean ± s.d.) were obtained from three replicate measurements.

Next we investigated the accelerated ATPase activity observed in $KaiC_{RS}$ after the formation of the complex. The ATPase activity reported for $KaiC_{SE}$ is low (about 15 ATP molecules per day per molecule of $KaiC_{SE}$) and was proposed as a reason for the slowness of circadian oscillation[28]. $KaiC_{RS}$ alone shows a significantly faster ATPase rate than $KaiC_{SE}$, which is further enhanced by binding of $KaiB_{RS}$ (208 ± 19 and 1,557 ± 172 ATP molecules day per day per $KaiC_{RS}$, respectively; left two bars in Fig. 3d and Extended Data Fig. 6c–g). Furthermore, $KaiC_{RS}$ does not exhibit temperature compensation for its ATPase activity (temperature coefficient Q_{10} about 1.9; Extended Data Fig. 6c), a feature that is present in $KaiC_{SE}$ and proposed to be a prerequisite for self-sustained rhythms[28]. The deviation from unity for Q_{10} is consistent with our earlier observation that the $KaiBC_{RS}$ system is not a true circadian oscillator but rather an hourglass timer (Fig. 3b).

Regulatory role of $KaiB_{RS}$

Mechanistic details of how the binding of $KaiB_{RS}$ in the CI domain allosterically affects the autodephosphorylation of $KaiC_{RS}$ in the CII domain remain unclear. There are three plausible scenarios to explain this: (1) $KaiB_{RS}$ binding stimulates the phosphoryl transfer from pSer back to ADP (Extended Data Fig. 7a); (2) $KaiB_{RS}$ binding increases the hydrolysis rate of the active-site ATP (Extended Data Fig. 8a); or (3) $KaiB_{RS}$ binding accelerates nucleotide exchange in the CII domain (Extended Data Fig. 8e). To differentiate among these possibilities, we performed radioactivity experiments to follow nucleotide interconversion. We also measured ATPase activity for wild-type $KaiC_{RS}$ and mutant forms that are incapable of ATPase

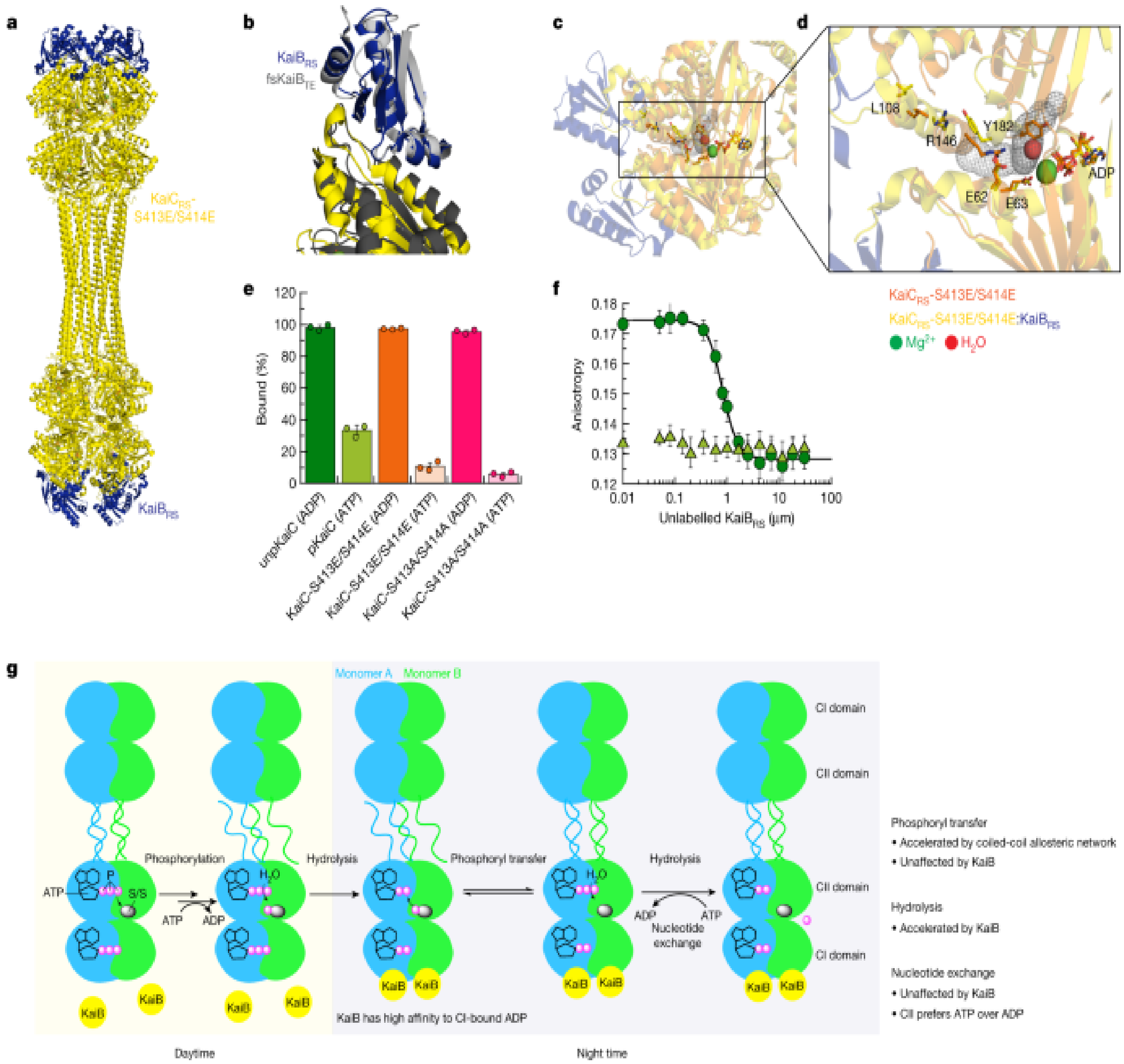
activity in the CI or CII domain, and quantified nucleotide-exchange rates by measuring the fluorescence of mant-ATP. First, we detected fast, transient ^{32}P-ATP formation in our radioactivity experiments when starting from ^{32}P-phosphorylated $KaiC_{RS}$, which was due to its ATP synthase activity in the CII domain (Fig. 3e and Extended Data Fig. 7b–d). The observed phosphoryl-transfer rate was independent of $KaiB_{RS}$ (observed rate constant = $12.0 \pm 1.7\ h^{-1}$ and $15.4 \pm 1.7\ h^{-1}$ in its absence and presence, respectively; Fig. 3f) and agreed well with the rates determined from our gel electrophoresis experiments ($11.0 \pm 0.8\ h^{-1}$ and $11.5 \pm 0.8\ h^{-1}$ with or without $KaiB_{RS}$, respectively; Extended Data Fig. 7e,f). Our experimental data confirmed that $KaiC_{RS}$ undergoes dephosphorylation through an ATP synthase mechanism, similar to what was observed for $KaiC_{SE}$ (ref. [29]). KaiB does not expedite the actual phosphoryl-transfer reaction, which is never the rate-limiting step. As we were unable to stabilize the first phosphorylation site (Ser414) in the presence of ADP, the rates reported here correspond exclusively to dephosphorylation of Ser413. Second, to deconvolute the contributions of the CI and CII domains to the observed ATPase activity, we measured ADP production from $KaiC_{RS}$ mutants that abolish hydrolysis in either the CI domain ($KaiC_{RS}$-E62Q/E63Q) or the CII domain ($KaiC_{RS}$-E302Q/E303Q). For wild-type $KaiC_{RS}$, the binding of $KaiB_{RS}$ resulted in a 7.5-fold increase in ATPase activity, and both domains were affected and contributed additively (3-fold for CI and at least 1.7-fold for CII) to the overall effect (Fig. 3d and Extended Data Fig. 8b–d). Of note, the fold increase in the CII domain represents a lower limit as the mutations induced to generate $KaiC_{RS}$-E62Q/E63Q interfere with $KaiB_{RS}$ binding, as previously reported for $KaiC_{SE}$ (ref. [30]). Third, our measurements of nucleotide exchange showed that this rate is also unaffected by $KaiB_{RS}$ binding ($19.8 \pm 1.8\ h^{-1}$ and $18.0 \pm 1.5\ h^{-1}$ with or without $KaiB_{RS}$, respectively; Fig. 3g). As there is no tryptophan residue near the nucleotide-binding site in the CI domain, only the exchange rate in the CII domain could be determined. Notably, the change in fluorescence amplitude was smaller in the presence of $KaiB_{RS}$, which demonstrates that even though the binding of $KaiB_{RS}$ does not accelerate nucleotide exchange, it appears to

induce a conformational rearrangement in the CII domain, especially at higher temperatures (Fig. 3g and Extended Data Fig. 8f–h).

Structure of the KaiBC$_{RS}$ complex

To elucidate the structural underpinning of the enhanced ATPase activity of KaiC$_{RS}$ after KaiB$_{RS}$ binding, we solved the cryo-EM structures of KaiC$_{RS}$ alone (PDB: 8FWI) and in complex with KaiB$_{RS}$ (PDB: 8FWJ) (Extended Data Table 2). Twelve KaiB$_{RS}$ molecules (monomeric in solution; Extended Data Fig. 9a) bind to the CI domain of the KaiC$_{RS}$-S413E/S414E dodecamer (Fig. 4a–c and Extended Data Fig. 9b). The bound state of KaiB$_{RS}$ adopts the same fold-switch conformation as observed for KaiB$_{TE}$ (ref. [25]) and suggests that this is the canonical binding-competent state (Fig. 4b). Following binding of KaiB$_{RS}$, the CI–CI interfaces loosen up (Fig. 4c), which enables the formation of a tunnel that connects bulk solvent to the position of the hydrolytic water in the active sites (Fig. 4d and Extended Data Fig. 9c). There are other lines of evidence for the weakened interactions within the CI domains. First, KaiB$_{RS}$ binding to either KaiC$_{RS}$-CI domain (Extended Data Fig. 10a) or KaiC$_{RS}$-Δcoil (that is, missing the C-terminal extensions; Extended Data Fig. 10b) resulted in disassembly of the hexameric KaiC$_{RS}$ structure into its monomers. By contrast, full-length KaiC$_{RS}$ maintained its oligomeric state following binding of KaiB$_{RS}$, which is probably due to the stabilization provided by the coiled-coil interaction. Second, a decrease in melting temperature (T_m) of KaiC$_{RS}$ was observed with increasing KaiB$_{RS}$ concentration (Extended Data Fig. 10c). There was no interaction between neighbouring KaiB$_{RS}$ molecules within the complex (Extended Data Fig. 9b), which suggests that there is a non-cooperative assembly of KaiB$_{RS}$ to KaiC$_{RS}$. This result is contrary to what has been observed for KaiBC$_{SE}$ and KaiBC$_{TE}$ complexes[31,32].

Fig. 4: KaiB$_{RS}$ binds to the post-hydrolysis state and accelerates the ATPase activity of KaiC$_{RS}$.



a, Cryo-EM structure of KaiC$_{RS}$-S413E/S414E (yellow) in complex with KaiB$_{RS}$ (blue) (PDB: 8FWJ). **b**, Superposition of KaiC$_{RS}$-S413E/S414E (yellow) bound to KaiB$_{RS}$ (blue) (PDB: 8FWJ) and KaiC$_{TE}$-S413E (dark grey) bound to fsKaiB$_{TE}$ (fold-switch, binding-competent state of KaiB$_{TE}$; light grey) (PDB: 5JWQ)[26]. **c**,**d**, Binding of KaiB$_{RS}$ (blue) creates a tunnel (grey mesh) that enables water to reach the catalytic position (red sphere) for ATP hydrolysis in the CI domain. **e**, Binding of wild-type and mutant forms of KaiC$_{RS}$ to His-tagged KaiB$_{RS}$ in the presence of ADP or an ATP-recycling system at 25 °C. Bar graphs show mean ± s.d. from three replicates. **f**, Fluorescence anisotropy of unlabelled KaiB$_{RS}$ competitively displacing KaiB$_{RS}$–6IAF (where 6IAF is the fluorophore) from

unphosphorylated $KaiC_{RS}$ in the presence of ADP (dark green circles) and phosphorylated $KaiC_{RS}$ in the presence of the ATP-recycling system (light green triangles) at 30 °C. The average anisotropy and standard error were calculated from ten replicate measurements. **g**, Schematic of the uncovered mechanism of $KaiC_{RS}$ regulated by coiled-coil interactions and $KaiB_{RS}$ in the CI and CII domains.

Furthermore, we noted that KaiB-bound structures in phosphomimetic variants of $KaiC_{RS}$ (Fig. 4c,d) and $KaiC_{SE}$ (ref. [26]) have ADP bound in their CI domain. This result demonstrates that the post-hydrolysis state is also the binding-competent state for $KaiB_{RS}$. To test this hypothesis, a His-tagged $KaiB_{RS}$ protein was used in pull-down assays to detect its physical interaction with wild-type and mutant forms of $KaiC_{RS}$ bound with either ADP or ATP. Nearly all $KaiB_{RS}$ was complexed to ADP-bound $KaiC_{RS}$, whereas less than 30% co-eluted in the ATP-bound form, regardless of the phosphorylation state (Fig. 4e and Extended Data Fig. 10d,e). The formation of complexes depended inversely on the ATP-to-ADP ratio (Extended Data Fig. 10f). We performed fluorescence anisotropy competition experiments to obtain a more quantitative description of the binding interaction between $KaiC_{RS}$ and $KaiB_{RS}$. Highly similar dissociation constant (K_d) values were obtained for unphosphorylated, wild-type $KaiC_{RS}$ (Fig. 4f) and its phosphomimetic form (Extended Data Fig. 10g) bound with ADP (0.42 ± 0.03 μM and 0.79 ± 0.06 μM, respectively). No measurable binding curves were obtained for ATP-bound phosphorylated wild-type $KaiC_{RS}$ (Fig. 4f) or for $KaiC_{RS}$-S413E/S414E (Extended Data Fig. 10g) with ATP-recycling system, which is probably due to the small fraction of complex present. Our data show that the post-hydrolysis state in the CI domain is key for $KaiB_{RS}$ binding, whereas the phosphorylation state of $KaiC_{RS}$ has only a marginal effect.

In summary, we demonstrate that binding of $KaiB_{RS}$ at the CI domain in the post-hydrolysis state facilitates the hydrolysis of transiently formed ATP after dephosphorylation of $KaiC_{RS}$ in the CII domain (Fig. 4g). Our fluorescence experiments (Fig. 3g and Extended Data Fig. 8f) detected a conformational change in the CII domain following $KaiB_{RS}$ binding, but we

did not observe major structural changes in the cryo-EM structures. Based on the temperature dependence of the fluorescence amplitudes (Extended Data Fig. 8f), we conjecture that the inability to detect conformational differences is probably because of the low temperature. As the CII domain prefers to bind ATP over ADP (Extended Data Fig. 10h), ATP hydrolysis in the CII domain stimulated by $KaiB_{RS}$ is particularly important to keep $KaiC_{RS}$ in its dephosphorylated state at night time. During this period, the exogenous ATP-to-ADP ratio remains sufficiently high to otherwise result in ATP-binding in the CII active site (Fig. 3c and Extended Data Fig. 6b).

Discussion

The $KaiBC_{RS}$ system studied here represents a primordial, hourglass timekeeping machinery, and its mechanism provides insight into more evolved circadian oscillators such as KaiABC. The dodecameric $KaiC_{RS}$ showed constitutive kinase activity owing to its extended C-terminal tail that forms a coiled-coil bundle with the opposing hexamer. This structure elicits a conformation akin to the exposed A loop conformation in $KaiAC_{SE}$, and autophosphorylation occurs within half an hour. In the $KaiABC_{SE}$ system, the transition from unphosphorylated to double phosphorylated KaiC takes place over about 12 h, and the fine-tuning of this first half of the circadian rhythm is accomplished by the emergence of $KaiA_{SE}$ during evolution. The second clock protein, KaiB, binds the CI domain with the same fold-switched state in both systems. The interaction is controlled by the phosphorylation state in the $KaiABC_{SE}$ system, and its sole function is to sequester $KaiA_{SE}$ from the activating binding site, whereas KaiB binding directly accelerates ATPase activity in the $KaiBC_{RS}$ system regardless of the phosphorylation state. The $KaiBC_{RS}$ system requires an environmental switch in the ATP-to-ADP concentration to reset the clock. The system therefore follows the day–night schedule when nucleotide concentrations inherently fluctuate in the organism. By contrast, the self-sustained oscillator $KaiABC_{SE}$ remains functional over a wide range of nucleotide concentrations and responds to changes in the ATP-to-ADP ratio by changing its phosphorylation period and amplitude to remain entrained with the day–night cycle[33].

The newly reported structural fold of KaiC utilizes the versatile coiled-coil architecture as part of a long-range allosteric network that regulates $KaiC_{RS}$ dephosphorylation. Nature uses conformational changes in coiled-coil domains for a variety of regulatory functions[34], including the activity of the motor protein dynein in the cellular transport of cargo along the actin filament[12]. A similar register shift, although in a coiled-coil interaction formed by only two helices, is used in dynein motility. Given that this simple heptad repeat sequence emerged multiple times and is found throughout all kingdoms of life[35], it is an example of convergent evolution.

Methods

Construct of KaiC and KaiB expression vectors

The wild-type $KaiC_{RS}$ (GenBank identifier: ACM04290.1) and $KaiB_{RS}$ (GenBank: WP_002725098.1) from *R. sphaeroides* strain KD131/KCTC 12085 (equivalent: *Cereibacter sphaeroides* strain KD131) constructs used in this paper were ordered from GenScript (Supplementary Table 1). Codon-optimized plasmids for $KaiC_{RS}$ and $KaiB_{RS}$ were subcloned into NcoI/KpnI sites of the pETM-41 vector. A QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies) was used to generate single mutant, double mutant and truncated versions of $KaiC_{RS}$. The truncated $KaiC_{RS}$ ($KaiC_{RS}$-Δcoil and $KaiCI_{RS}$) were generated by introducing stop codons in the $KaiC_{RS}$ wild-type plasmid. All primers were ordered from Genewiz (Supplementary Table 2). The presence of the intended $KaiC_{RS}$ mutations in the plasmid was confirmed by DNA sequencing by Genewiz using primers ordered from the same company (listed in Supplementary Table 2).

Both $KaiC_{SE}$ and $KaiA_{SE}$ plasmids were a gift from E. K. O'Shea. Expression and purification were performed according to a previously described procedure[6].

Expression and purification of $KaiC_{RS}$ and $KaiB_{RS}$ from *R. sphaeroides*

$KaiC_{RS}$, $KaiC_{RS}$ mutants and $KaiB_{RS}$ were expressed in *Escherichia coli* BL21(DE3) cells (New England Biolabs) harbouring the plasmid pETM-41 containing the $kaiC_{RS}$ or $kaiB_{RS}$ gene. Three colonies from a freshly prepared transformation plate were inoculated into 1 litre of TB medium containing 50 μg ml^{-1} kanamycin. This culture was grown at 25 °C with shaking at 220 r.p.m. for 48 h without IPTG induction (leaky expression). The cells were pelleted by centrifugation at 4,200 r.p.m. for 15 min at 4 °C and stored at −80 °C.

Frozen cell pellets of $KaiC_{RS}$ and $KaiC_{RS}$ mutants were resuspended into lysate buffer (buffer $A_{C\text{-}RS}$) containing 1× EDTA-free protease inhibitor cocktail (Thermo Fisher Scientific), DNAse I (Sigma Aldrich) and lysozyme (Sigma Aldrich), and the lysate was sonicated for 10–15 min (20 s on, 30 s off, output power less than 40%) on ice followed by centrifugation at 18,000 r.p.m. at 4 °C for 45 min to remove cell debris. The lysate was filtered through a 0.45 μm filter and then loaded on HisTrap HP prepacked Ni-sepharose columns (Cytiva) pre-equilibrated with buffer $A_{C\text{-}RS}$ at 0.5 ml min^{-1}. The column was washed with buffer $A_{C\text{-}RS}$ at 1 ml min^{-1} until the UV absorbance returned to baseline. Impurities were then washed with 15% buffer $B_{C\text{-}RS}$, and the protein was eluted with 50% buffer $B_{C\text{-}RS}$. The eluted complex was diluted with 1.5-fold dialysis $buffer_{C\text{-}RS}$ then subjected to in-house prepared His-tagged TEV protease (1:10, TEVP:$KaiC_{RS}$ molar ratio) cleavage to remove the His_6–MBP tag from $KaiC_{RS}$ (wild-type and mutants) overnight at 4 °C in 6–8 kDa snakeskin dialysis tubing (Thermo Fisher Scientific) that was exchanged against dialysis $buffer_{C\text{-}RS}$. Cleaved $KaiC_{RS}$ was filtered through a 1 μm filter and once again loaded onto HisTrap HP prepacked Ni-sepharose columns at 0.5 ml min^{-1} to remove His-tagged TEV protease, His_6–MBP tag and uncleaved protein. The flow through was concentrated using a Millipore Amicon Ultra-15 centrifugal filter device (10 kDa cut-off) and immediately passed through a HiPrep Sephacryl S-400 HR column (Cytiva) pre-equilibrated with buffer $C_{C\text{-}RS}$. Protein was purified to homogeneity with a single band on Bis-Tris 4–12% gradient SDS–PAGE gel (Genscript) at 62.5 kDa. All protein purification steps were done at 4 °C or on ice. Protein was aliquoted and flash-frozen before storage at −80 °C until further use. The protein concentration was measured using a Microplate

BCA Protein Assay kit (Thermo Fisher Scientific) on a SpectraMax MiniMax 300 imaging cytometer using BSA as a standard curve. Typical yields of $KaiC_{RS}$ (wild-type and mutants) were 20–40 mg l^{-1} of culture.

To test whether the purified $KaiC_{RS}$ had any ATPase contamination, Q-sepharose HP columns (Cytiva) pre-equilibrated with buffer $D_{C\text{-}RS}$ was used before the final HiPrep Sephacryl S-400 HR column. The protein was eluted with 5 CV of a linear gradient from 0 to 100% buffer $E_{C\text{-}RS}$. The ATPase activity of the protein samples purified using Q-sepharose HP columns was identical to the samples without this additional purification step. Buffer $A_{C\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 250 mM NaCl, 10 mM imidazole, 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v). Buffer $B_{C\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 250 mM NaCl, 500 mM imidazole, 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v). $Dialysis_{C\text{-}RS}$ comprised 50 mM Tris-base (pH 7.0), 50 mM NaCl, 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v). Buffer $C_{C\text{-}RS}$ comprised 50 mM Tris-base (pH 7.0), 50 mM NaCl, 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v). Buffer $D_{C\text{-}RS}$ comprised50 mM Tris-base (pH 7.0), 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v). Buffer $E_{C\text{-}RS}$ comprised 50 mM Tris-base (pH 7.0), 1 M NaCl, 2 mM TCEP, 5 mM $MgCl_2$, 1 mM ATP and 10% glycerol (v/v).

The purification of $KaiB_{RS}$ was similar to $KaiC_{RS}$, but with slight modifications as outlined below. After sonication and centrifugation to remove cell debris, the lysate was filtered through a 0.22 µm filter and then passed through HisTrap HP prepacked Ni Sepharose columns, pre-equilibrated with buffer $A_{B\text{-}RS}$. The column was washed with buffer $A_{B\text{-}RS}$ until the UV absorbance returned to baseline. Impurities were washed with 5% buffer $B_{B\text{-}RS}$, and the protein was eluted with 50% buffer $B_{B\text{-}RS}$. The fusion protein was concentrated down to around 30 ml using Amicon stirred cells (Millipore Sigma) with 10 kDa cut-off. In-house prepared His-tagged TEV protease was added, and the fusion protein was cleaved overnight at 4 °C in a 3.5 kDa dialysis cassette that was exchanged against $dialysis_{B\text{-}RS}$ buffer. The cleaved $KaiB_{RS}$ was passed through HisTrap HP prepacked Ni-

sepharose columns and concentrated down to about 10 ml using a Millipore Amicon Ultra-15 centrifugal filter device (3.5 kDa cut-off). The protein sample was then loaded onto a 26/60 Superdex S75 gel-filtration column (Cytiva) pre-equilibrated with buffer $C_{B\text{-}RS}$ at 4 °C. The eluted protein was loaded onto Q-sepharose HP columns pre-equilibrated with buffer $D_{B\text{-}RS}$ to remove ATPase contamination. The protein was eluted out in the flow through. A gradient from 0 to 100% buffer $E_{B\text{-}RS}$ was passed through Q-sepharose HP columns to ensure that no $KaiB_{RS}$ was bound to the columns. Protein was purified to homogeneity with the single band on Bis-Tris 4–12% gradient SDS–PAGE gels at 10.3 kDa. Protein was aliquoted and flash-frozen before storage at −80 °C until use. The protein concentration was measured using a Microplate BCA Protein Assay kit (Thermo Fisher Scientific) on a SpectraMax MiniMax 300 imaging cytometer using BSA as a standard curve. Typical yields of $KaiB_{RS}$ were 30–40 mg l^{-1} of culture.

Buffer $A_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 250 mM NaCl, 10 mM imidazole, 2 mM TCEP and 10% glycerol (v/v). Buffer $B_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 250 mM NaCl, 500 mM imidazole, 2 mM TCEP and 10% glycerol (v/v). $Dialysis_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 250 mM NaCl, 2 mM TCEP and 10% glycerol (v/v). Buffer $C_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 50 mM NaCl, 2 mM TCEP and 10% glycerol (v/v). Buffer $D_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 2 mM TCEP and 10% glycerol (v/v). Buffer $E_{B\text{-}RS}$ comprised 50 mM Tris-base (pH 7.5), 1 M NaCl, 2 mM TCEP and 10% glycerol (v/v).

Phylogenetic tree of KaiC

Protein sequences used in this study were identified in a multistep process. In the first step, a selection of sequences was identified using the BLASTP algorithm, utilizing a query based on the protein sequence for KaiC from *S. elongatus* (GenBank: WP_011242648.1)[38]. The query was run against NCBI's non-redundant protein database with the exclusion of models or uncultured and environmental sample sequences. A multiple sequence alignment of the selected 1,538 sequences was generated using MAFFT[39,40,41] (Supplementary Dataset 1). This alignment was used as input

to generate an initial phylogenetic tree for KaiC with RAxML (v.8.2.9)[42] using the PROTGAMMALG model. The generated tree was then used to identify the emergence of KaiB to create the final tree that focused on systems containing either KaiBC or KaiABC. To do so, a BLASTP search was performed for each branch tip with the KaiB sequence from *S. elongatus*, with the results restricted to the organism at which the branch tip was identified from. The observed branch point of emergence of KaiB agrees well with previous results in which it was shown that KaiB is mainly seen in non-archaea, non-proteobacteria[2].

To ensure the best possible sequence coverage, a BLASTP search using KaiB as query (GenBank: WP_011242647.1) was performed. The resulting sequences were then used to identify the organisms that they came from, which allowed us to create a list of organisms with an identified KaiB sequence. This list was then used to select for a subsequent BLASTP search using KaiC as query and therefore to identify only KaiC sequences for organisms that contain both KaiB and KaiC. A spot check was run to confirm that, for example, KaiA was indeed found among all cyanobacteria identified except for *Prochlorococcus marinus*. The obtained sequences were trimmed down to only include sequences with a sequence homology of 90% or less using CD-HIT[43] to arrive at a total of 401 sequences. For the calculation of the phylogenetic tree, RecA from *S. elongates* was added to serve as the outgroup. Sequences were aligned using MAFFT[39,40,41] (Supplementary Dataset 2) with the E-INS-I algorithm[44]. The multiple sequence alignment was then used as input for the phylogenetic tree calculation with IQ-TREE (v.1.6.beta5), using the LG-substitution matrix[45] with the freeRate model (using 10 categories; LG+R10)[46,47]. To enable determination of branch support, an aBayes test[48], a SH-aLRT test (20,000 bootstrap replicates[49]) and an ultrafast bootstrap (20,000 bootstrap replicates[50]) were performed (Supplementary Dataset 1; branch supports in order: SH-aLRT support (%)/aBayes support/ultrafast bootstrap support (%)).

X-ray crystallography

$KaiC_{RS}$ and $KaiC_{RS}$-Δcoil crystals were obtained by sitting-drop vapour diffusion using a 96-well Intelli-Plate (102-0001-00, Art Robbins) at 291 K. Drops contained 0.5 µl crystallization solution, 0.5 µl protein at 10 mg ml^{-1} in 20 mM MOPS pH 6.5, 50 mM NaCl, 2 mM TCEP, 5 mM $MgCl_2$, 3 mM ATP and 1 mM AMPPCP, and were equilibrated against 50 µl of the solution in the reservoir. The $KaiC_{RS}$ crystallization solution consisted of 200 mM magnesium chloride hexahydrate, 100 mM HEPES pH 7.5 and 30% (w/v) PEG 400. $KaiC_{RS}$-Δcoil crystals were grown using 200 mM ammonium acetate, 100 mM sodium citrate tribasic dihydrate pH 5.6 and 30% (w/v) PEG 4,000.

The PEG 400 in the $KaiC_{RS}$ crystallization solution acted as a cryoprotectant, whereas $KaiC_{RS}$-Δcoil crystals were cryoprotected in LV Cryo Oil (MiTeGen). Single crystals were cooled in liquid nitrogen, and X-ray diffraction images were collected at ALS beamline 8.2.1 at 100 K (data collection details are described in Extended Data Table 1). The data were indexed and integrated in iMosflm[51], and scaled and merged in Aimless[52].

To obtain a structural model of $KaiC_{RS}$, first the $KaiC_{RS}$-Δcoil structure was solved by molecular replacement in MRage[53] using the $KaiC_{RS}$ sequence (residues 1–490) as input to search for homologues in the PDB database. The initial $KaiC_{RS}$-Δcoil structure based on the $KaiC_{SE}$ (PDB: 1TF7 (ref. [36])) was manually rebuilt in Coot (v.0.9.81)[54] and refined in Phenix (v.1.20.1-4487)[55]. Finally, the $KaiC_{RS}$-Δcoil structure was used as the molecular replacement search model in Phaser[56] to solve the full-length $KaiC_{RS}$ structure.

The assigned space groups were validated in Zanuda[57], and the position of the asymmetric unit in the unit cell was standardized using Achesym[58]. $KaiC_{RS}$ coiled-coil registers were analysed using SamCC-Turbo (v.0.0.2) with the default socket cut-off value of 7.4 (ref. [59]). The images of protein structures were rendered using PyMOL (v.2.6.0)[60].

Tunnel detection and calculation were performed using CAVER 3.0.2 PyMOL plugin[61], with the minimum probe radius varying between 0.9 and

1.1. Default values were used for all other parameters. All atoms except waters were used in the calculation. The residue selection for starting point consisted of Glu62, Glu63 and ADP602. The catalytic position of the water in the CI domain was modelled from the crystal structure of the transition-state analogue-bound F_1-ATPase (PDB: 1w0j, water 2,064 from chain D[62]).

Cryo-EM and image processing

For preparation of EM grids, 3–4 μl of 4.3 mg ml^{-1} (per monomer concentration) of sample in 20 mM MOPS pH 6.50, 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 2 mM ATP was applied to glow-discharged 1.2/1.3 400 mesh C-flat carbon-coated copper grids (Protochips). The grids were frozen using a Vitrobot Mark IV (ThermoFisher) at 4 °C and 95% humidity, with a blot time of 4 s. All datasets were collected on a Titan Krios operated at an acceleration voltage of 300 keV, with a GIF quantum energy filter (Gatan) and a GATAN K2 Summit direct electron detector controlled by SerialEM[63].

Inspection of the raw cryo-EM images revealed some heterogeneity in the relative orientations between individual hexamers of the dodecameric particles, presumably due to inherent flexibility in the coiled-coil regions, which limited the resolution to 3.3–3.4 Å. To obtain higher resolution reconstructions, the dodecamers were split and processed as individual hexamers, with *C*6 symmetry being applied throughout processing. To reconstitute the full dodecamer reconstruction, two copies of the hexamer reconstruction were overlaid on top of each other using the 'fit in map' function in Chimera[64] to fit one hexamer into the lower resolution end density of the other. The overlaid hexamers were then combined, creating a new map in which each voxel takes the value from the hexamer with highest absolute value.

For $KaiC_{RS}$-S413E/S414E alone, a dataset of approximately 2,500 movies was collected. The movies were recorded with a pixel size of 1.074 Å, including 70 frames and with an exposure rate of 1.31 e^- per $Å^2$ per frame. Approximately 825,000 particles were picked, and after 2D classification, around 320,000 particles from good class averages were carried forward for

further processing. The final measured resolution of the reconstruction was 2.9 Å (Extended Data Fig. 4a).

For the $KaiC_{RS}$-S413E/S414E:$KaiB_{RS}$ complex, a dataset of around 2,000 movies was collected. The movies were recorded with a pixel size of 1.023 Å, including 70 frames and with an exposure rate of 1.35 e^{-} per $Å^2$ per frame. About 440,000 particles were picked, and after 2D classification around 190,000 particles from good class averages were carried forward for further processing. The final measured resolution of the reconstruction was 2.7 Å (Extended Data Fig. 4b).

All data processing was carried out using *cis*TEM (v.2.0.0)[65], and followed the workflow of motion correction, CTF parameter estimation, particle picking, 2D classification, ab initio 3D map generation, 3D refinement, 3D classification, per-particle CTF refinement and B-factor sharpening. The highest resolution of 3D refinement used was 4 Å for both reconstructions, and final resolutions were estimated using the *cis*TEM PartFSC and a threshold of 0.143.

To validate the combined dodecamer structure, we also processed both datasets as full *D*6 symmetric dodecamers (Extended Data Fig. 4c,d). This was accomplished by extracting the picked hexamers into a large box size and performing 2D classification with automatic centring. Clear dodecamer class averages were then selected and re-extracted from the original images, with picking coordinates that were adjusted by the translation required to match the centred class average. After this centring, duplicate picks were removed to obtain the final dodecamer particle stacks. These stacks were processed as described above, with the highest resolution of 3D refinement used as 4.25 Å.

In an attempt to find deviations from *D*6 symmetry, we also calculated reconstructions for both structures assuming *C*1 symmetry, starting from the ab initio 3D step. The resulting refined *C*1 structures did not exhibit detectable departures from D6 symmetry (Extended Data Fig. 4e). We therefore present symmetrized volumes as our final result.

The cryo-EM structures were built using the $KaiC_{RS}$ model obtained by X-ray crystallography and fold-switch-stabilized $KaiB_{TE}$ (PDB: 5JWO (ref. [26])) as starting points. The models were constructed using Coot (v.0.9.81)[54], and refinement was carried out using Phenix (v.1.20.1-4487)[55].

Preparation of unphosphorylated $KaiC_{RS}$

Purified $KaiC_{RS}$ (about 20 µM) from −80 °C was dialysed in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 0.1 mM ADP overnight at 4 °C to remove glycerol and to replace ATP with ADP. The dialysed $KaiC_{RS}$ was then heated at 30 °C for 4 h to obtain fully unphosphorylated $KaiC_{RS}$ bound with ADP, and the sample was then passed through 0.22 µm Spin-X centrifuge tube filters (Corning). The sample was concentrated to a higher concentration (less than 100 µM) at 4 °C. The protein concentration was measured using a BCA assay.

In vitro $KaiBC_{RS}$ reaction

Kinetics of $KaiC_{RS}$ autophosphorylation in the presence and absence of $KaiB_{RS}$

Unphosphorylated $KaiC_{RS}$ (3.5 µM, prepared as described above) in the presence or absence of $KaiB_{RS}$ (3.5 µM and 35 µM) was preincubated at 20, 25, 30 and 35 °C for 1 h in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 0.1 mM ADP. The reactions were started by adding 3.9 mM ATP to obtain a final concentration of 4 mM nucleotide in the presence of 2 U ml^{-1} pyruvate kinase (Millipore Sigma) and 10 mM phosphoenolpyruvate (Millipore Sigma) to regenerate ATP during the reaction. The reaction samples were sampled by hand at specific time points and mixed with an equal amount of loading dye (stock concentration of 0.1 M Tris-base (pH 7.5), 4% SDS, 0.2% bromophenol blue, 30% glycerol and 0.5 M 2-mercaptoethanol). The mixed samples were then stored at −20 °C until further use.

Kinetics of $KaiC_{RS}$ autodephosphorylation in the presence and absence of $KaiB_{RS}$

Purified $KaiC_{RS}$ (around 20 μM) was dialysed in reaction buffer containing 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 0.1 mM ADP overnight at 4 °C to remove glycerol and to generate $KaiC_{RS}$ bound with ADP. After dialysis at 4 °C, $KaiC_{RS}$ exists in two states: 50% unphosphorylated and 50% single phosphorylated at Ser413 (pSer413), which were confirmed by tandem mass spectrometry (data not shown). The autodephosphorylation reaction was started by adding $KaiC_{RS}$ or $KaiC_{RS}$ in the presence of $KaiB_{RS}$ (3.5 μM) into reaction buffer pre-equilibrated at 30 °C.

Oscillation of $KaiBC_{RS}$

Dialysed $KaiC_{RS}$ (3.5 μM) was preincubated at 35 °C for 30 min in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 0.1 mM ADP in the presence or absence of $KaiB_{RS}$ (3.5 μM). The reactions were started by adding 4 mM ATP and reaction samples were collected at specific time points for 10% SDS–PAGE and HPLC analysis to identify phosphorylation state of $KaiC_{RS}$ and amount of nucleotide at each time point, respectively.

Controlling ATP-to-ADP ratio to mimic daytime and night time

$KaiC_{RS}$ was dialysed in reaction buffer containing 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 1 mM ATP overnight at 4 °C. To start the reaction as shown in Fig. 3c, $KaiC_{RS}$ (3.5 μM) in the absence or presence of $KaiB_{RS}$ (3.5 μM) was mixed with additional ATP (final 4 mM to mimic daytime), and the reaction samples (500 μl) were added into a D-Tube Dialyzer (midi 3.5 kDa cut-off, EMD Millipore) that was exchanged against 4 mM ATP buffer (400 ml). After the 12-h time point, the reaction samples were transferred into preincubated 25% ATP/ADP buffer (400 ml) that

mimics the night time. After the 24-h time point, the same samples were changed into preincubated 4 mM ATP to mimic the daytime again.

To start the experiment as shown in Extended Data Fig. 6a, $KaiC_{RS}$ (35 μM) in the presence of 3 mM ATP in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ was heated at 35 °C for 25 min to generate fully phosphorylated $KaiC_{RS}$. The $KaiC_{RS}$ sample was then diluted 10-fold into 25% ATP/ADP buffer pre-equilibrated at 30 °C to final concentration of $KaiC_{RS}$ (3.5 μM) and $KaiB_{RS}$ (3.5 μM or 35 μM). The reaction samples (300 μl) were added into a D-Tube Dialyzer (midi 3.5 kDa cut-off, EMD Millipore) that was exchanged against 25% ATP/ADP buffer (300 ml).

During the reaction, the samples were gently shaken in a 30 °C incubator, and reaction samples were collected at specific time points for 10% SDS–PAGE and HPLC analysis to identify the phosphorylation state of $KaiC_{RS}$ and the amount of nucleotide at each time point, respectively.

The rationale for the ATP-to-ADP ratio at daytime and night time comes from two earlier literature reports. The change in ATP-to-ADP ratio at daytime and night time were directly measured in vivo in the strain *R. sphaeroides*[27], in which ATP is 2.0–2.4 mM during day and drops to 0.5–0.6 during night, and it is well known that the total nucleotide concentration stays constant. We chose the total nucleotide concentration of 4 mM in our in vitro work to be identical to the described in vitro experiments performed for the canonical $KaiC_{SE}$. Because of photosynthesis in daylight, virtually all nucleotide is ATP[33]. We note that a slightly higher amount of ATP will not affect our results, as the affinity of $KaiC_{RS}$ for ATP is higher than for ADP.

Separation of unphosphorylated, single and double phosphorylated $KaiC_{RS}$ by SDS–PAGE

Unphosphorylated, single phosphorylated and double phosphorylated $KaiC_{RS}$ were separated by 10% SDS–PAGE with 37.5:1 acrylamide:bis-acrylamide (Bio-Rad), 18 cm × 16 cm × 1 mm Tris-HCl gel with 1× Tris-glycine SDS running buffer (Invitrogen). The samples were heated at 95 °C for 3 min, and 400 ng of material was loaded onto the Tris-HCl gel. The gel

was run with a constant current of 35 mA, 150 W, and the voltage was greater than 700 V for 5.5 h in a cold room, with a water bath set to 12 °C using a Hoefer SE600 electrophoresis unit.

Unphosphorylated and phosphorylated $KaiC_{RS}$-Δcoil were separated by Zn^{2+} Phos-tag SDS–PAGE with 10% acrylamide gel containing 50 µM Phos-tag acrylamide (Wako). The gel was run with a constant current of 30 mA for 5 h 30 min in a cold-room, with 1 µg per well protein samples pre-heated at 95 °C for 3 min.

The gels were stained overnight at room temperature with InstantBlue protein gel stain (Expedeon) with gentle shaking and destained with distilled water until bands were clearly visible. The gels were imaged on a ChemiDoc Imager (Bio-Rad), and Image Lab software (Bio-Rad) was used for analysis.

Statistics and reproducibility for gel electrophoresis

Data shown in main text figures and Extended Data figures are representative SDS–PAGE gels for at least three independent biological replicates ($n = 3$), except for experiments presented in Fig. 3a,c, which were performed in duplicate.

Oligomerization state of $KaiC_{RS}$ and $KaiB_{RS}$

Gel-filtration chromatography

$KaiC_{RS}$, unphosphorylated $KaiC_{RS}$ and all $KaiC_{RS}$ mutants with a concentration of around 40–80 µM were loaded with a flow rate of 0.2 ml min^{-1} onto a prepacked Superdex-200 10/300 GL (GE Healthcare) pre-equilibrated with 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 1 mM ATP (0.1 mM ADP for unphosphorylated $KaiC_{RS}$) at 4 °C using an ÄKTA Pure system (GE Healthcare). $KaiB_{RS}$ (0.5 mM) was loaded onto Superdex-75 10/300 GL (GE Healthcare) pre-equilibrated with 20 mM MOPS (pH 6.5), 50 mM NaCl and 2 mM TCEP at 4 °C. The eluate was collected in fractions of 1 ml each and subjected to SDS–PAGE analysis. A standard curve (that is, molecular weight versus elution time)

was determined for the column using molecular weight protein standards (Bio-Rad) run in the same buffer and flow rate. The protein standard mixture contained thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B_{12} (1.35 kDa).

Analytical ultracentrifugation

Sedimentation velocity centrifugation experiments were run at 50,000 r.p.m. (for $KaiB_{RS}$) and 30,000 r.p.m. (for $KaiC_{RS}$ wild-type and mutants and $KaiC_{SE}$), with continuous scans from 5.8 to 7.3 cm at 0.005 cm intervals at 20 °C on a Beckman Optima XL-A (Beckman-Coulter) equipped with absorption optics and a four-hole An60Ti rotor. Measurements were set up at 280 nm (for $KaiB_{RS}$) and 295 nm (for $KaiC_{RS}$ and $KaiC_{SE}$) to avoid interference from ATP. The software package SEDFIT (v.14.1) was used for data evaluation[66]. $KaiC_{RS}$ and $KaiC_{RS}$-Δcoil (100 μM) were prepared in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 1 mM ATP. $KaiC_{SE}$ (100 μM) was prepared in 20 mM MOPS (pH 8.0), 150 mM NaCl, 2 mM TCEP, 5 mM $MgCl_2$, and 1 mM ATP. $KaiB_{RS}$ (500 μM) was prepared in 20 mM MOPS (pH 6.5), 50 mM NaCl and 2 mM TCEP.

ATPase activity

Purified $KaiC_{RS}$ (both wild-type and mutant forms (around 20 μM)) and $KaiB_{RS}$ (about 90 μM) were dialysed in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 1 mM ATP (reaction buffer) overnight at 4 °C. The samples were passed through 0.22 μm Spin-X centrifuge tube filters, and concentrations were measured using a BCA assay before setting up the reactions. Typical $KaiC_{RS}$ or $KaiBC_{RS}$ reactions contained 3.5 μM $KaiC_{RS}$ (wild type and mutants) and 3.5 μM $KaiB_{RS}$ in reaction buffer with a final concentration of 4 mM ATP. The samples were incubated at the indicated temperatures and were sampled by hand at specific time points. Next 10 μl of sample was quenched with 10 μl of 10% trichloroacetic acid (Millipore Sigma), and the mixture was passed through a 0.22 μm Spin-X centrifuge tube filter to remove the precipitated protein. The flow through sample was then re-adjusted to pH 6.2 for nucleotide

separation by adding 10 µl of 0.75 M HEPES, pH 8.0. The final samples were kept at −20 °C until HPLC analysis.

Three microlitres of each sample were injected with a high-precision autosampler (injection error of <0.1 µl, resulting in a maximum systemic error of about 6%) to a reverse-phase HPLC instrument with an ACE 5 µm particle size, C18-AR and 100 Å pore size column (Advanced Chromatography Technologies). The instrument was pre-equilibrated with 100 mM potassium phosphate pH 6.2 with a flow rate of 0.4 ml min^{-1}. Using pure nucleotide samples, the retention times of ATP, ADP and AMP were determined to be 2.6, 3.1 and 4.4 min, respectively. The concentration of each nucleotide was calculated from the relative ratio of the peak areas and the total nucleotide concentration. To determine ATPase activity rates, the observed rate constants were determined from at least five data points for each temperature using initial rate analysis and least-squares linear regression (Extended Data Figs. 6c,d and 8b–d). The mean values and uncertainties (s.d.) shown in Fig. 3d, Extended Data Figs. 6 and 8 were derived from three replicate experiments. KaleidaGraph (v.4.5.3; Synergy) was used for data analysis and plotting.

Nucleotide exchange

$KaiC_{RS}$ (wild type or mutants) and $KaiB_{RS}$ were dialysed into 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 50 µM ATP overnight at 4 °C. The samples were passed through 0.22 µm Spin-X centrifuge tube filters, and the protein concentration was measured using a BCA assay. The reaction contained 3.5 µM of $KaiC_{RS}$-S413E and/or 35 µM of $KaiB_{RS}$, and samples were incubated at 20, 25, 30 and 35 °C for 16–24 h in the presence of an ATP-recycling system. The reactions were started by adding 250 µM of mant-ATP (Jena Bioscience). The spectrum was measured using the fluorescence energy transfer from tryptophan residues in $KaiC_{RS}$ to mant-ATP by exciting the sample at 290 nm (2.5 nm bandwidth) and collecting the emission intensity from 320 nm to 550 nm (5 nm bandwidth) in increments of 2 nm. To measure the nucleotide exchange rate, the maximum change in fluorescence intensity at 440 nm ($\Delta F_{440\ nm}$) was followed for a total time of 1,800 s in 15 s increments with anti-

photobleaching mode on FluoroMax-4 spectrofluorometer (Horiba Scientific) equipped with a water bath to control the temperature. There are two tryptophan residues within 5 Å from the nucleotide-binding site in the $KaiCII_{RS}$ domain and no tryptophan residue close to the nucleotide-binding site in the $KaiCI_{RS}$ domain, so the nucleotide exchange observed in the experiments are for the $KaiCII_{RS}$ domain. To ensure that the exchange rate observed in the experiments are from nucleotide exchange in the $KaiCII_{RS}$ domain, $KaiCI_{RS}$ (which only contains the CI domain) was tested; no change in fluorescence was observed following the addition of mant-ATP (Extended Data Fig. 8h).

The experiments for $KaiC_{SE}$ alone and with $KaiC_{SE}$ mixed in and for $KaiA_{SE}$ were performed in a similar way, except that $KaiC_{SE}$ and $KaiA_{SE}$ were dialysed in 20 mM MOPS (pH 8.0), 150 mM NaCl, 2mM TCEP, 10 mM $MgCl_2$ and 50 µM ATP overnight at 4 °C. $KaiA_{SE}$ was incubated with $KaiC_{SE}$ for 1 h at 30 °C before adding 250 µM mant-ATP.

For the nucleotide preference experiment, $KaiC_{RS}$-S413E/S414E and $KaiC_{RS}$-S413A/S414A were dialysed in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 20 µM ADP overnight at 4 °C. $KaiC_{RS}$-S413E/S414E or $KaiC_{RS}$-S413A/S414A (3.5 µM) was first mixed with mant-ATPγS or mant-ADP (150 µM), and the kinetic trace at 440 nm was recorded at 30 °C. After the fluorescence trace at 440 nm reached a plateau, which indicates that the nucleotide analogue was fully bound to the protein, a 27-fold excess of ATP (4 mM) was added to displace the bound nucleotide analogue, and the decay of fluorescence intensity was recorded at 440 nm at 30 °C. The experiments were run in triplicate, and results were averaged and fitted to a single exponential decay.

Analysis was performed by fitting individual traces to an exponential equation using KinTek Explorer software[67,68]], and error bars denote the standard errors as obtained from triplicate experiments. KaleidaGraph (v.4.5.3; Synergy) was used for data plotting.

^{32}P-ATP radioactive labelling and experiment

^{32}P-labelled $KaiC_{RS}$ was prepared by mixing unphosphorylated $KaiC_{RS}$ (10 μM) with 0.46 μM [γ-^{32}P]ATP (3,000 Ci $mmol^{-1}$, PerkinElmer) and 500 μM ATP in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ at 35 °C for 30 min and then immediately switched to 4 °C to prevent dephosphorylation of $KaiC_{RS}$. The ^{32}P-labelled $KaiC_{RS}$ sample was passed through Zeba spin desalting columns (Thermo Fisher Scientific) pre-equilibrated with 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ twice at 4 °C. The sample was incubated with buffer containing 1 mM ADP overnight at 4 °C to obtain ^{32}P-labelled $KaiC_{RS}$ bound with ADP. The sample was passed through a final Zeba desalting column pre-equilibrated with 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ at 4 °C, and the solution was then incubated with 8 mM ADP in the presence or absence of $KaiB_{RS}$ at 4 °C for 1 h. The samples were then diluted in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ preincubated at 30 °C to obtain a final concentration of ^{32}P-labelled $KaiC_{RS}$ (5 μM), $KaiB_{RS}$ (20 μM) and ADP (4 mM). The reactions were incubated at 30 °C, and at different time points, aliquots were taken (1.5 μl). The reactions were stopped by adding 1.5 μl Laemmli sample buffer (62.5 mM Tris (pH 6.8), 2% SDS, 25% glycerol and 0.01% bromophenol blue) supplemented with 5% (v/v) 2-mercaptoethanol.

The samples were spotted onto a TLC plate (PEI-cellulose F plates, Merck) and quickly dried with a blow-dryer for 30 s. The TLC plates were run first with distilled water as a mobile phase. After TLC plates were completely dried, 0.75 M KH_2PO_4 was used as the mobile phase to separate ^{32}P-labelled $KaiC_{RS}$, [γ-^{32}P]ATP and inorganic phosphate (^{32}P), as previously shown[29]. The phosphor-screens were scanned on an Amersham Typhoon (GE Healthcare) at a resolution of 100 μm. ImageQuant TL 7.0 software was used for analysis.

Fluorescence anisotropy competition

Fluorescence anisotropy competition experiments were carried out using a FluoroMax-4 spectrofluorometer (Horiba Scientific) at 30 °C. Excitation and emission wavelengths for $KaiB_{RS}$ labelled with 6-iodoacetamidofluorescein (6-IAF, Thermo Fisher Scientific) at Cys29 were set at 492 nm (5 nm bandwidth) and 520 nm (5 nm bandwidth), respectively, with fixed G factor (G factor of $KaiB_{RS}$_6IAF alone) to eliminate instrumental bias. The average anisotropy and standard error were calculated from ten replicate measurements.

$KaiB_{RS}$–6IAF was prepared by mixing degassed $KaiB_{RS}$ (100 µM) with 20-fold excess of 6-IAF (stock 10 mM in 50% DMSO) in 20 mM Tris-base (pH 7.0), 50 mM NaCl and 1 mM TCEP (degassed). The reaction was incubated at room temperature for 4 h and dialysed against 20 mM Tris-base (pH 7.0), 50 mM NaCl and 1 mM TCEP in a 3.5 kDa dialysis cassette overnight at 4 °C to remove unreacted 6-IAF and small amounts of DMSO. The crosslinked sample was passed through a 0.22 µm Spin-X centrifuge tube filter and loaded onto Superdex-75 10/300 GL pre-equilibrated with 20 mM MOPS (pH 7.0), 50 mM NaCl and 1 mM TCEP with a 0.2 ml min^{-1} flow rate to remove leftover unreacted 6-IAF. The sample was aliquoted and flash-frozen in liquid nitrogen and stored at −80 °C until use. All the crosslinked reactions were performed in the dark.

For the fluorescence anisotropy competitive binding experiment, $KaiB_{RS}$–6IAF (0.2–0.4 µM) was first incubated with wild-type or mutant $KaiC_{RS}$ (1 µM, 60% increase in anisotropy in comparison to $KaiB_{RS}$–6IAF alone) in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ in the presence of 4 mM ADP or 4 mM ATP with an ATP-recycling system (2 U ml^{-1} pyruvate kinase and 10 mM phosphoenolpyruvate), then unlabelled $KaiB_{RS}$ (0–50 µM) was added to the samples. The samples were incubated at 30 °C for 4 h or 12 h before measurement.

The decrease in fluorescence anisotropy (FA) versus concentration of unlabelled $KaiB_{RS}$ was fitted to equation (1) using the Levenberg–

Marquardt nonlinear fitting algorithm included in KaleidaGraph (Synergy Software) to obtain the half-maximum inhibitory concentration (IC_{50}) value. The K_d value can then be calculated from the IC_{50} value using equation (2) as previously described[69].

$${\rm{FA}}={m}_{1}+\frac{({m}_{2}-{m}_{1})}{1+{10}^{{\rm{\log }}x-{\rm{\log }}({{\rm{IC}}}_{50})}}$$
(1)

$${K}_{{\rm{d}}}=\frac{{{\rm{I}}{\rm{C}}}_{50}}{1+\frac{[{\rm{l}}{\rm{a}}{\rm{b}}{\rm{e}}{\rm{l}}{\rm{l}}{\rm{e}}{\rm{d}}\,{{\rm{K}}{\rm{a}}{\rm{i}}{\rm{B}}}_{{\rm{R}}{\rm{S}}}]}{{K}_{{\rm{d}}}}}$$
(2)

Pull-down assays

Pull-down assays probe the interaction between $KaiC_{RS}$ (wild type and mutants) with $KaiB_{RS}$ (His_6-MBP-TEV-$KaiB_{RS}$). The His_6-MBP-TEV-$KaiB_{RS}$ was expressed and purified as described above but without the TEV protease cleavage step. Wild-type or mutant $KaiC_{RS}$ (3.5 µM) was mixed with $KaiB_{RS}$-tag (3.5 µM) in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP and 10 mM $MgCl_2$ in the presence of 4 mM ADP or 4 mM ATP with an ATP-recycling system in a final volume of 400 µl. The samples were incubated at 25 °C for 4 h or 24 h before loading onto a 500 µl spin column with 200 µl (prepared from 400 µl of 50% slurry) Talon beads (Takara) pre-equilibrated with sample buffer. The samples were incubated with the Talon beads for 30 min with gentle shaking, after which the flow through was collected by gravity into 1 ml Eppendorf tubes. The beads were washed three times with 400 µl sample buffer by gravity, then the samples were eluted with 200 µl of 0.5 M imidazole buffer by centrifugation at 1,000*g* for 1 min. The protein mixture, flow through, wash and eluted samples were run on a Bis-Tris 4–12% gradient SDS–PAGE gel with a molecular weight marker. The gels were stained with Coomassie blue and were imaged on a ChemiDoc Imager (Bio-Rad). Image Lab Software (Bio-Rad) was used for analysis.

The following control experiments were performed: (1) fusion $KaiB_{RS}$ protein in the absence of $KaiC_{RS}$; and (2) $KaiC_{RS}$ in the absence of fusion $KaiB_{RS}$. All the fusion $KaiB_{RS}$ proteins came out only in the elution buffer in the first control experiment, which indicated that fusion $KaiB_{RS}$ binds to Talon beads and the amount of $KaiB_{RS}$ used did not overload the column. All $KaiC_{RS}$ protein came out in the flow through in the second control experiment, which indicated there is no specific binding between $KaiC_{RS}$ and the Talon beads.

Thermofluor assay

$KaiC_{RS}$ and SYPRO Orange (Thermo Fisher Scientific) were used at final concentration of 3 µM and 10×, respectively. The experiments were carried out in 20 mM MOPS (pH 6.5), 50 mM NaCl, 2 mM TCEP, 10 mM $MgCl_2$ and 4 mM ADP or ATP. The samples were prepared to a final volume of 20 µl in a MicroAmp Fast Optical 96-well reaction plate (Applied Biosystems Life Technologies), and the plate was sealed with Axygen UltraClear sealing film (Corning). The assay plate was run in a StepOne Real-Time PCR instrument (Applied Biosystems Life Technologies) with melt curve set up. The temperature was continuously increased from 25 °C to 95 °C by 0.3 °C every 15 s. The data were fit with nonlinear fitting in KaleidaGraph (Synergy Software) to a Boltzmann sigmoidal curve (equation (3)).

$$Y={\rm{Bottom}}+\frac{({\rm{Top}}-{\rm{Bottom}})}{1+{{\rm{\exp }}}^{\left(\frac{{T}_{{\rm{m}}}-T}{c}\right)}}$$
(3)

where Y is the fluorescence intensity at temperature T, T is the temperature in degrees Celsius, Bottom is the baseline fluorescence at low temperature, Top is the maximum fluorescence at the top of the truncated data, c is the slope or steepness of the curve, and T_m is the melting temperature of the protein.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Structure factors and refined models obtained using X-ray crystallography have been deposited into PDB under accession codes 8DBA (wild-type $KaiC_{RS}$) and 8DB3 ($KaiC_{RS}$-Δcoil). Cryo-EM maps and refined models have been deposited into the Electron Microscopy Data Bank (EMDB) and PDB, respectively. The composite map and model for the $KaiC_{RS}$-S413E/S414E dodecamer reconstruction are submitted under entries EMD-29505 and 8FWI, respectively. The composite map and model for the KaiCRS-S413E/S414E–$KaiB_{RS}$ dodecamer reconstruction are submitted under entries EMD-29506 and 8FWJ, respectively. The focused $KaiC_{RS}$-S413E/S414E hexamer refinement map is available under accession EMD-29507 and the focused $KaiC_{RS}$-S413E/S414E–$KaiB_{RS}$ hexamer refinement map is available under accession EMD-29508. The full $KaiC_{RS}$-S413E/S414E dodecamer refinement is available under accession EMD-29509 and the full $KaiC_{RS}$-S413E/S414E–$KaiB_{RS}$ dodecamer refinement is available under accession EMD-29510. Other datasets used are all publicly available in public community or discipline-specific repositories (for example, PDB identifiers 5JWQ, 1W0J, 1TF7 and 7S65). The accession codes for protein sequences, sequence alignments and phylogeny are listed in Supplementary Datasets 1 and 2.

References

1. Ishiura, M. et al. Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* **281**, 1519–1523 (1998).

2. Dvornyk, V., Vinogradova, O. & Nevo, E. Origin and evolution of circadian clock genes in prokaryotes. *Proc. Natl Acad. Sci. USA* **100**, 2495–2500 (2003).

3. Cohen, S. E. & Golden, S. S. Circadian rhythms in cyanobacteria. *Microbiol. Mol. Biol. Rev.* **79**, 373–385 (2015).

4. Golden, S. S. Principles of rhythmicity emerging from cyanobacteria. *Eur. J. Neurosci.* **51**, 13–18 (2020).

5. Partch, C. L. Orchestration of circadian timing by macromolecular protein assemblies. *J. Mol. Biol.* **432**, 3426–3448 (2020).

6. Rust, M. J., Markson, J. S., Lane, W. S., Fisher, D. S. & O'Shea, E. K. Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* **318**, 809–812 (2007).

7. Holtzendorff, J. et al. Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *J. Biol. Rhythms* **23**, 187–199 (2008).

8. Axmann, I. M. et al. Biochemical evidence for a timing mechanism in *Prochlorococcus*. *J. Bacteriol.* **191**, 5342–5347 (2009).

9. Ma, P., Mori, T., Zhao, C., Thiel, T. & Johnson, C. H. Evolution of KaiC-dependent timekeepers: a proto-circadian timing mechanism confers adaptive fitness in the purple bacterium *Rhodopseudomonas palustris*. *PLoS Genet.* **12**, e1005922 (2016).

10. Dvornyk, V. & Knudsen, B. Functional divergence of the circadian clock proteins in prokaryotes. *Genetica* **124**, 247–254 (2005).

11. Min, H., Guo, H. & Xiong, J. Rhythmic gene expression in a purple photosynthetic bacterium, *Rhodobacter sphaeroides*. *FEBS Lett.* **579**, 808–812 (2005).

12. Carter, A. P. et al. Structure and functional role of dynein's microtubule-binding domain. *Science* **322**, 1691–1695 (2008).

13. Schmelling, N. M. et al. Minimal tool set for a prokaryotic circadian clock. *BMC Evol. Biol.* **17**, 169 (2017).

14. Aoki, S. & Onai, K. in *Bacterial Circadian Programs* (eds Mackey S. R. Ditty J. L. & Johnson C. H.) 259–282 (Springer, 2009).

15. Wiegard, A. et al. Biochemical analysis of three putative KaiC clock proteins from *Synechocystis* sp PCC 6803 suggests their functional divergence. *Microbiology* **159**, 948–958 (2013).

16. Pattanayek, R. et al. Analysis of KaiA–KaiC protein interactions in the cyano-bacterial circadian clock using hybrid structural methods. *EMBO J.* **25**, 2017–2028 (2006).

17. Nishiwaki-Ohkawa, T., Kitayama, Y., Ochiai, E. & Kondo, T. Exchange of ADP with ATP in the CII ATPase domain promotes autophosphorylation of cyanobacterial clock protein KaiC. *Proc. Natl Acad. Sci. USA* **111**, 4455–4460 (2014).

18. Kim, Y. I., Dong, G., Carruthers, C. W. Jr., Golden, S. S. & LiWang, A. The day/night switch in KaiC, a central oscillator component of the circadian clock of cyanobacteria. *Proc. Natl Acad. Sci. USA* **105**, 12825–12830 (2008).

19. Swan, J. A. et al. Coupling of distant ATPase domains in the circadian clock protein KaiC. *Nat. Struct. Mol. Biol.* **29**, 759–766 (2022).

20. Egli, M. et al. Loop–loop interactions regulate KaiA-stimulated KaiC phosphorylation in the cyanobacterial KaiABC circadian clock. *Biochemistry* **52**, 1208–1220 (2013).

21. Hayashi, F. et al. ATP-induced hexameric ring structure of the cyanobacterial circadian clock protein KaiC. *Genes Cells* **8**, 287–296 (2003).

22. Nishiwaki, T. et al. Role of KaiC phosphorylation in the circadian clock system of *Synechococcus elongatus* PCC 7942. *Proc. Natl Acad. Sci. USA* **101**, 13927–13932 (2004).

23. Xu, Y. et al. Identification of key phosphorylation sites in the circadian clock protein KaiC by crystallographic and mutagenetic analyses.

Proc. Natl Acad. Sci. USA **101**, 13933–13938 (2004).

24. Kitayama, Y., Iwasaki, H., Nishiwaki, T. & Kondo, T. KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *EMBO J.* **22**, 2127–2134 (2003).

25. Chang, Y. G. et al. Circadian rhythms. A protein fold switch joins the circadian oscillator to clock output in cyanobacteria. *Science* **349**, 324–328 (2015).

26. Tseng, R. et al. Structural basis of the day–night transition in a bacterial circadian clock. *Science* **355**, 1174–1180 (2017).

27. Abee, T., Hellingwerf, K. J. & Konings, W. N. Effects of potassium ions on proton motive force in *Rhodobacter sphaeroides*. *J. Bacteriol.* **170**, 5647–5653 (1988).

28. Terauchi, K. et al. ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc. Natl Acad. Sci. USA* **104**, 16377–16381 (2007).

29. Nishiwaki, T. & Kondo, T. Circadian autodephosphorylation of cyanobacterial clock protein KaiC occurs via formation of ATP as intermediate. *J. Biol. Chem.* **287**, 18030–18035 (2012).

30. Phong, C., Markson, J. S., Wilhoite, C. M. & Rust, M. J. Robust and tunable circadian rhythms from differentially sensitive catalytic domains. *Proc. Natl Acad. Sci. USA* **110**, 1124–1129 (2013).

31. Murakami, R. et al. Cooperative binding of KaiB to the KaiC hexamer ensures accurate circadian clock oscillation in cyanobacteria. *Int. J. Mol. Sci.* **20**, 4550 (2019).

32. Snijder, J. et al. Insight into cyanobacterial circadian timing from structural details of the KaiB–KaiC interaction. *Proc. Natl Acad. Sci. USA* **111**, 1379–1384 (2014).

33. Rust, M. J., Golden, S. S. & O'Shea, E. K. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331**, 220–223 (2011).

34. Truebestein, L. & Leonard, T. A. Coiled-coils: the long and short of it. *Bioessays* **38**, 903–916 (2016).

35. Liu, J. & Rost, B. Comparing function and structure between entire proteomes. *Protein Sci.* **10**, 1970–1979 (2001).

36. Pattanayek, R. et al. Visualizing a circadian clock protein: crystal structure of KaiC and functional insights. *Mol. Cell* **15**, 375–388 (2004).

37. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

39. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

41. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**, 1899–1900 (2010).

42. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

43. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

44. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).

45. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320 (2008).

46. Soubrier, J. et al. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **29**, 3345–3358 (2012).

47. Yang, Z. A space–time process model for the evolution of DNA sequences. *Genetics* **139**, 993–1005 (1995).

48. Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C. & Gascuel, O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* **60**, 685–699 (2011).

49. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

50. Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

51. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281 (2011).

52. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).

53. Bunkoczi, G. et al. Phaser.MRage: automated molecular replacement. *Acta Crystallogr. D* **69**, 2276–2286 (2013).

54. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

55. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).

56. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

57. Lebedev, A. A. & Isupov, M. N. Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda. *Acta Crystallogr. D* **70**, 2430–2443 (2014).

58. Kowiel, M., Jaskolski, M. & Dauter, Z. ACHESYM: an algorithm and server for standardized placement of macromolecular models in the unit cell. *Acta Crystallogr. D* **70**, 3290–3298 (2014).

59. Szczepaniak, K., Bukala, A., da Silva Neto, A. M., Ludwiczak, J. & Dunin-Horkawicz, S. A library of coiled-coil domains: from regular bundles to peculiar twists. *Bioinformatics* **36**, 5368–5376 (2020).

60. The PyMOL Molecular Graphics System, version 2.4 (Schrödinger, 2020).

61. Chovancova, E. et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).

62. Kagawa, R., Montgomery, M. G., Braig, K., Leslie, A. G. & Walker, J. E. The structure of bovine F1-ATPase inhibited by ADP and beryllium fluoride. *EMBO J.* **23**, 2734–2744 (2004).

63. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

64. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

65. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* https://doi.org/10.7554/eLife.35383 (2018).

66. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).

67. Johnson, K. A. Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol.* **467**, 601–626 (2009).

68. Johnson, K. A., Simpson, Z. B. & Blom, T. Global Kinetic Explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387**, 20–29 (2009).

69. Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K_I) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).

Acknowledgements

D.K. and N.G. are supported by the Howard Hughes Medical Institute (HHMI). We would like to thank M. Rigney for assistance with negative-stain data collection at the Brandeis University Electron Microscopy Facility, and Z. Yu and staff at the Janelia Research Campus cryo-EM facility for advice and assistance with data collection. The Berkeley Center for Structural Biology is supported in part by HHMI. Beamline 8.2.1 of the Advanced Light Source, a US DOE Office of Science User Facility under contract no. DE-AC02-05CH11231, is supported in part by the ALS-ENABLE programme funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01. Mass

spectral data were obtained at the University of Massachusetts Mass Spectrometry Core Facility (RRID:SCR_019063).

Author information

Author notes

1. Warintra Pitsawong

   Present address: Biomolecular Discovery, Relay Therapeutics, Cambridge, MA, USA

2. Timothy Grant

   Present address: John and Jeanne Rowe Center for Research in Virology, Morgridge Institute for Research, Madison, Madison, WI, USA

3. Timothy Grant

   Present address: Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

4. Marc Hoemberger & Renee Otten

   Present address: Treeline Biosciences, Watertown, MA, USA

5. Nikolaus Grigorieff

   Present address: Howard Hughes Medical Institute, RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA

6. These authors contributed equally: Warintra Pitsawong, Ricardo A. P. Pádua, Timothy Grant

Authors and Affiliations

1. Howard Hughes Medical Institute and Department of Biochemistry, Brandeis University, Waltham, MA, USA

   Warintra Pitsawong, Ricardo A. P. Pádua, Marc Hoemberger, Renee Otten & Dorothee Kern

2. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

   Timothy Grant & Nikolaus Grigorieff

3. Department of Biochemistry, Brandeis University, Waltham, MA, USA

   Niels Bradshaw

Contributions

W.P., R.A.P.P. and D.K. conceived the project and designed experiments. W.P. performed and analysed all biochemical data. W.P. and R.A.P.P. set up the crystal trays. R.A.P.P. collected and analysed the X-ray crystallographic data. W.P. prepared the samples for the cryo-EM studies and collected negative-stain images to screen for optimal sample conditions. T.G. collected and processed all cryo-EM data and reconstructed the cryo-EM maps under supervision of N.G. R.A.P.P. built and interpreted the structural models. W.P. and N.B. performed and analysed experiments with radioactively labelled KaiC. M.H. built the KaiC phylogeny tree. W.P., R.O. and D.K. wrote the paper. All authors commented on the manuscript and contributed to data interpretation.

Corresponding author

Correspondence to Dorothee Kern.

Ethics declarations

Competing interests

D.K. is co-founder of Relay Therapeutics and MOMA Therapeutics. The remaining authors declare no competing interests.

Peer review

Peer review information

Nature thanks Marta Carroni, Satchidananda Panda and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Evolution of *kaiC* and sequence alignment of *kaiC* subgroups.

(a) Phylogenetic tree of *kaiC* homologs, where *kaiC* genes that have an approximately 50 amino acids C-terminal extension are labeled in red. *Rhodobacter sphaeroides* strain KD131 studied here and *Synechococcus elongatus* PCC 7942 (widely studied in the literature) are highlighted in green and pink, respectively. The accession code and organism are shown at the tip of the branches, the numbers at each node represent the aBayes bootstrap values[48], and the legend for branch length is shown (see also Supplementary Datasets 1 and 2). **(b)** A sequence alignment of the CII domain of the *kaiC* subgroups annotated with its sequence similarity. Residue Glu490, the position where the stop codon was introduced in the truncated $KaiC_{RS}$-Δcoil construct, is shown in red and marked with an arrow.

Extended Data Fig. 2 Auto-phosphorylation and nucleotide exchange rates of KaiC.

(a) 10% SDS-PAGE gel of 3.5 μM $KaiC_{RS}$ in the presence of 4 mM ATP and using an ATP-recycling system at 30 °C. U, S, and D represent unphosphorylated, singly, and doubly phosphorylated $KaiC_{RS}$, respectively. **(b)** Densitometric analysis of auto-phosphorylation (single + double phosphorylation) from panel (a) over time yields a rate of 6.5 ± 1.0 h^{-1}. **(c)** 6.5% SDS-PAGE gel of 3.5 μM $KaiC_{SE}$ in the presence of 1.2 μM $KaiA_{SE}$ and 4 mM ATP at 30 °C. U and P represent unphosphorylated and phosphorylated $KaiC_{SE}$, respectively. **(d)** Densitometric analysis of auto-phosphorylation of $KaiC_{SE}$ activated by $KaiA_{SE}$ (panel (c)) shows a rate of 0.40 ± 0.02 h^{-1} and is substantially slower than for $KaiC_{RS}$. The standard deviation for parameters in (b) and (d) were obtained from data fitting. **(e)** 10% SDS-PAGE gels for experiments with 3.5 μM $KaiC_{RS}$ in the presence of 4 mM ATP and using an ATP-recycling system between 20 and 35 °C show that the level of phosphorylation increases with temperature. U, S, and D represent unphosphorylated, singly, and doubly phosphorylated $KaiC_{RS}$, respectively. For gel source data in (a), (c), and (e), see Supplementary Figure 2. **(f)** Bar graphs indicating the nucleotide exchange rate in the CII domain of $KaiC_{RS}$ incubated with 50 μM ATP in the presence of an ATP-recycling system, and then mixed with 250 μM mant-ATP. An increase in fluorescence intensity at 440 nm was recorded and the single-exponential time traces were fitted to obtain the exchange rate constants: 3.6 ± 0.8 h^{-1} (20 °C), 12.2 ± 1.0 h^{-1} (25 °C), 18.5 ± 1.5 h^{-1} (30 °C), and 25.2 ± 0.2 h^{-1} (35 °C). Experiments were performed in triplicate and data are presented as mean values ± s.d.

Extended Data Fig. 3 Oligomeric states of $KaiC_{RS}$ and effect of coiled-coil domain on rates of nucleotide exchange and auto-phosphorylation.

(a) Oligomerization analysis of $KaiC_{RS}$ (dodecamer, green line) and truncated $KaiC_{RS}$-Δcoil (hexamer, cyan line) by analytical gel-filtration

chromatography. The protein size markers are indicated at the top. **(b)** Comparison of the elution profiles of $KaiC_{RS}$-Δcoil (cyan line) and $KaiC_{SE}$ (purple line) from size-exclusion chromatography shows a hexameric state for both $KaiC_{RS}$-Δcoil and $KaiC_{SE}$. **(c)** Oligomeric states of $KaiC_{RS}$ (dodecamer, green line), $KaiC_{RS}$-Δcoil (hexamer, cyan line), and $KaiC_{SE}$ (hexamer, purple line) were also measured by analytical ultracentrifugation (sedimentation velocity at 30,000 rpm and 20 °C) and the results agree with the data shown in panels (a) and (b). The graph in panel (c) represents the sedimentation coefficient distribution [c(s)]. **(d)** The change in fluorescence at 440 nm (ΔF_{440nm}) represents the nucleotide exchange between ATP and mant-ATP at 30 °C for $KaiC_{RS}$ (green trace, $18.0 \pm 1.5\ h^{-1}$) and $KaiC_{RS}$-Δcoil (cyan trace, $19.1 \pm 0.8\ h^{-1}$). Representative traces are shown and the fitted parameters (mean ± s.d.) were obtained from three replicate measurements. **(e)** Zn^{2+} Phos-tag™ SDS-PAGE gel shows the level of phosphorylation over time of $KaiC_{RS}$ (upper gel) and $KaiC_{RS}$-Δcoil (lower gel) at 35 °C. P and U represent phosphorylated and unphosphorylated protein, respectively. **(f)** Phosphorylation level over time of $KaiC_{RS}$ (green circles, $7.4 \pm 0.3\ h^{-1}$) and $KaiC_{RS}$-Δcoil (cyan circles, $5.5 \pm 0.4\ h^{-1}$) analyzed by densitometric analysis of Zn^{2+} Phos-tag™ SDS-PAGE gel in (e). **(g)** First derivative of thermal-stability curves measured for unphosphorylated $KaiC_{RS}$ bound with ADP (brown line) and phosphorylated $KaiC_{RS}$ bound with ATP (green line). The extracted temperatures of denaturation are 50 °C (unphosphorylated $KaiC_{RS}$ in the presence of 1 mM ADP) and 58 °C (phosphorylated $KaiC_{RS}$ in the presence of 1 mM ATP), respectively. **(h)** Dodecameric state of unphosphorylated $KaiC_{RS}$ (40 μM) bound with ADP measured by size-exclusion chromatography. **(i)** SDS-PAGE gel shows dephosphorylation of Ser413 over time at 30 °C in the presence of 4 mM ADP (U and pSer413 represent unphosphorylated and Ser413-phosphorylated $KaiC_{RS}$, respectively) with the corresponding kinetics shown in the right panel (confirmed by MS/MS) with a rate constant of $11.5 \pm 0.8\ h^{-1}$. This result suggests that the coiled-coil domain promotes $KaiC_{RS}$ dephosphorylation. For gel source data in (e) and (i), see Supplementary Figure 2. The standard deviation for parameters in (f) and (i) were obtained from data fitting.

Extended Data Fig. 4 Graphical description of the cryo-EM processing workflow and validation of the final dodecamer structures.

The workflow (see description in Methods section) demonstrates a typical image (scale bar: 60 nm) and representative good class averages. The ab initio and final reconstructions are shown. Shown alongside the final reconstruction is the angular plot demonstrating the distribution of particle views and the Fourier shell correlation curve used for the global resolution estimation **(a)** $KaiC_{RS}$-S413E/S414E alone and **(b)** $KaiC_{RS}$-S413E/S414E:$KaiB_{RS}$ complex. To validate the final combined dodecamer structures, the data were reprocessed for the full dodecamer. The figure shows representative good class averages, the final reconstruction, angular distribution and Fourier shell correlation curve for the **(c)** $KaiC_{RS}$-S413E/S414E alone and **(d)** $KaiC_{RS}$-S413E/S414E:$KaiB_{RS}$ complex dodecamers. **(e)** Comparison of the *C*1 and *D*6 reconstructions of $KaiC_{RS}$-S413E/S414E alone and $KaiC_{RS}$-S413E/S414E:$KaiB_{RS}$, and Fourier shell correlation curves for the *C*1 reconstructions. The *C*1/*D*6 comparisons do not reveal discernable differences, suggesting that these complexes have *D*6 symmetry.

Extended Data Fig. 5 Correlation between the coiled-coil register shift and phosphorylation, and model for consecutive phosphorylation/dephosphorylation events in CII domain of $KaiC_{RS}$.

(a) Structural comparison between $KaiC_{RS}$ (green) and $KaiC_{RS}$-S413E/S414E (orange, single-chain for clarity) reveals that the coiled-coil in the phosphomimetic structure points outwards, with an angle of about 20° relative to the $KaiC_{RS}$ coiled-coil. **(b)** The conformational change in the coiled-coil domain affects the dimer interface due to partner swaps with the opposite hexamer (see also Fig. 2). From an “outside perspective”: the C-terminal helix in $KaiC_{RS}$ interacts with the right chain from the opposite hexamer, whereas in $KaiC_{RS}$-S413E/S414E the interaction is with the chain

on the left. **(c)** Coiled-coil diagrams describe the heptad register shift that accompanies this structural rearrangement. **(d)** Based on the overlay of our structures, we propose the following model for the phosphorylation/dephosphorylation events. First, the phosphorylation cycle starts with the transfer of the γ-phosphate of ATP to the hydroxyl group of Ser414 (1; green arrow) in unphosphorylated $KaiC_{RS}$ (green) or $KaiC_{RS}$-Δcoil (cyan). Secondly, pSer414 of $KaiC_{RS}$-Δcoil (purple, singly phosphorylated) moves away from the active site placing the hydroxyl group of Ser413 closer to the γ-phosphate of ATP for the second phosphorylation (2; purple arrow). Thirdly, the doubly phosphomimetic state ($KaiC_{RS}$-S413E/S414E, orange) reveals that the phosphoryl group of pSer414 moves back towards the active site for dephosphorylation (3; orange arrow). Lastly, we hypothesize that the indole group of Trp419 "pushes" pSer413 into the active site for the second dephosphorylation event (dashed arrow), in agreement with the slower dephosphorylation rate observed in the $KaiC_{RS}$-Δcoil construct (*cf*. Fig. 2d).

Extended Data Fig. 6 Effect of ATP-to-ADP ratio on $KaiC_{RS}$ auto-dephosphorylation and the dependence of temperature and $KaiB_{RS}$ binding on the ATPase activity of $KaiC_{RS}$.

(a) 10% SDS-PAGE gel of 3.5 µM $KaiC_{RS}$ and 3.5 µM $KaiB_{RS}$ in the presence of 4 mM ATP and 10 mM 2-phosphoenolpyruvate at 30 °C shows that the auto-phosphorylation cycle restarts upon regeneration of ATP by the addition of 2 U ml^{-1} pyruvate kinase at the 24-hour time mark. **(b)** 10% SDS-PAGE gel of 3.5 µM $KaiC_{RS}$ without (upper panel) and with (lower panel) 3.5 µM $KaiB_{RS}$ in the presence of 4 mM ATP with an ATP-recycling system added from the beginning showing that under these conditions KaiB does not accelerate dephosphorylation. For gel source data in (a) and (b), see Supplementary Figure 2. **(c)** Representative curves for ADP production of $KaiC_{RS}$ (3.5 µM) alone and **(d)** in the presence of $KaiB_{RS}$ (3.5 µM) in 4 mM ATP measured by HPLC. The data were analysed as described in the Methods section and result in ATPase activities of 108 ± 10 day^{-1} $KaiC^{-1}$ (with $KaiB_{RS}$ = 176 ± 29 day^{-1} $KaiC^{-1}$) at 20 °C, 163 ± 16 day^{-1} $KaiC^{-1}$

(with $KaiB_{RS}$ = 1052 ± 143 day^{-1} $KaiC^{-1}$) at 25 °C, 208 ± 19 day^{-1} $KaiC^{-1}$ (with $KaiB_{RS}$ = 1557 ± 172 day^{-1} $KaiC^{-1}$) at 30 °C, and 300 ± 21 day^{-1} $KaiC^{-1}$ (with $KaiB_{RS}$ = 2584 ± 245 day^{-1} $KaiC^{-1}$) at 35 °C. The temperature coefficient, Q_{10}, was calculated using the data obtained at 25 °C and 35 °C and yields a value of ~ 1.9. The standard deviations of ATPase activity at each temperature (right panels) were obtained from three replicate measurements and data are presented as mean values ± s.d. **(e)** The comparison of ADP production of $KaiC_{RS}$ in the absence (orange circles, data from panel (c)) and presence (orange diamonds, data from panel (d)) of $KaiB_{RS}$ at 30 °C indicate a 7.5-fold increase in ATPase activity for the complex. The binding of $KaiB_{RS}$ accelerates the ATPase activity of $KaiC_{RS}$ in both the CI and CII domains (see also Extended Data Fig. 8b, c). **(f)** The SDS-PAGE gel of $KaiC_{RS}$ (10 µg) and $KaiB_{RS}$ (10 µg) shows that both proteins were purified to homogeneity and the measured ATPase activity is, therefore, not due to impurities. **(g)** ADP production of $KaiB_{RS}$ in 4 mM ATP at 30 °C shows, as expected, no ATPase activity for $KaiB_{RS}$ alone and confirms the increase in ATPase activity shown in panel (d) is due to complex formation. The standard deviation for the representative curves shown in panels (c-d, left), e, and g was set to 6% assuming the largest systematic error originates from the injector.

Extended Data Fig. 7 Dephosphorylation of $KaiC_{RS}$ occurs via an ATP-synthase mechanism and the phosphoryl-transfer step is unaffected by the binding of $KaiB_{RS}$.

(a) Possible mechanisms how $KaiB_{RS}$ could accelerate $KaiC_{RS}$ dephosphorylation at nighttime. Binding of $KaiB_{RS}$ on the CI_{RS} domain directly accelerates the phosphoryl transfer from pSer to bound ADP to generate transiently bound ATP. The cartoon represents the interface between two monomers in the CII_{RS} domain. **(b)** Autoradiograph of separation of ^{32}P-$KaiC_{RS}$ at Ser413, transiently formed ^{32}P-ATP, and free ^{32}Pi via thin-layer chromatography (TLC) with 4 mM ADP at 30 °C, with the corresponding kinetics shown in **(c)** where gray circle, purple triangle,

and cyan diamonds represent the relative concentrations of phosphorylated ^{32}P-KaiC$_{RS}$, ^{32}P-ATP, and free ^{32}Pi, respectively. **(d)** Comparison of transient ^{32}P-ATP formation and decay in the absence (open triangle) and presence (solid triangle) of KaiB$_{RS}$ and free ^{32}P formation in the absence (open circles) and presence of KaiB$_{RS}$ (solid circles). Faster decay of transient ^{32}P-ATP together with higher free ^{32}P production in the presence of KaiB$_{RS}$ indicated that KaiB$_{RS}$ accelerates hydrolysis in KaiC$_{RS}$. **(e)** SDS-PAGE gel (10%) of dephosphorylation of phosphorylated 3.5 μM KaiC$_{RS}$ at Ser413 without (upper gel) and with (lower gel) 3.5 μM KaiB$_{RS}$ in the presence of 4 mM ADP at 30 °C (for gel source data, see Supplementary Figure 2). **(f)** Densitometric analysis of data in panel (e) shows the decay of total KaiC$_{RS}$ phosphorylation in the absence (gray circles) and presence (red diamonds) of KaiB$_{RS}$ and yields rates of 11.5 ± 0.8 h^{-1} and 11.0 ± 0.8 h^{-1}, respectively. This result indicates that binding of KaiB$_{RS}$ does not accelerate the phosphoryl-transfer step in KaiC$_{RS}$.

Extended Data Fig. 8 Effect of KaiB$_{RS}$ binding on ATPase activity and nucleotide exchange in the CII domain of KaiC$_{RS}$.

(a) Second possible mechanism to explain how KaiB$_{RS}$ accelerates KaiC$_{RS}$ dephosphorylation at nighttime**:** binding of KaiB$_{RS}$ to the CI$_{RS}$ domain could increase the hydrolysis rate in the CII$_{RS}$ domain and, thereby, prevent the phosphoryl transfer back from transiently formed or external ATP back to serine residues. **(b)** Representative curves for ADP production of phosphorylated KaiC$_{RS}$ with catalytic mutations in the CI domain (KaiC$_{RS}$-E62Q/E63Q, 3.5 μM) in the absence (dark green circles) and presence (light green circles) of 3.5 μM KaiB$_{RS}$ at 30 °C with 4 mM ATP was quantified using HPLC. From these data an ATPase activity in the CII domain of 112 ± 8 day^{-1} KaiC^{-1} and 195 ± 16 day^{-1} KaiC^{-1} in the presence of KaiB$_{RS}$ was determined. **(c)** Representative curves for ADP production measured by HPLC as in panel (b) of KaiC$_{RS}$ but with catalytic mutations in the CII domain (KaiC$_{RS}$-E302Q/E303Q, 3.5 μM) in the absence (dark pink circles)

and presence (light pink diamonds) of 3.5 μM $KaiB_{RS}$. The corresponding ATPase activities in the CI domain are 110 ± 12 day^{-1} $KaiC^{-1}$ in the absence and 320 ± 22 day^{-1} $KaiC^{-1}$ in the presence of $KaiB_{RS}$. **(d)** Representative curves for ADP production of $KaiCI_{RS}$-E62Q/E63Q (construct of only CI domain with catalytic mutations) in 4 mM ATP at 30 °C shows no ATPase activity indicating that Glu62 and Glu63 are the only two residues that are responsible for ATPase activity in CI domain of $KaiC_{RS}$ and confirms the ATPase activity shown in panel b is due to ATPase activity in CII domain of $KaiC_{RS}$. The standard deviation for the representative curves shown in panels (b-d) was set to 6% assuming the largest systematic error originates from the injector. The experiments in panel (b-d) were performed in triplicate and ATPase rate given in the legend for panels (b) and (c) are presented as mean values ± s.d. **(e)** Third possible mechanism to explain how $KaiB_{RS}$ accelerates $KaiC_{RS}$ dephosphorylation at nighttime: binding of $KaiB_{RS}$ to the CI_{RS} domain could promote faster nucleotide exchange in the CII_{RS} domain to displace transient ATP by ADP. **(f)** Time course of fluorescence intensity at 440 nm due to binding of mant-ATP to $KaiC_{RS}$-S413E bound with ATP in the absence (solid blue trace) and presence (red dotted trace) of $KaiB_{RS}$. $KaiC_{RS}$-S413E (3.5 μM) was pre-incubated with 3.5 μM $KaiB_{RS}$ for 16 h at 20, 25, 30, and 35 °C in the presence of 50 μM ATP and an ATP-recycling system and then mixed with 250 μM mant-ATP. The observed exchange rates at each temperature are listed in the table **(g)**. **(h)** Nucleotide exchange of $KaiCI_{RS}$ (i.e., only CI_{RS} domain) cannot be measured since there is no tryptophan residue in close proximity of the nucleotide binding site. In summary, $KaiB_{RS}$ accelerates $KaiC_{RS}$ dephosphorylation by increasing the hydrolysis rate in the CI and CII domains and does not affect the nucleotide exchange rate. Representative traces are shown in (f) and (h) and the fitted parameters (g; mean ± s.d.) were obtained from three replicate measurements.

Extended Data Fig. 9 KaiB-KaiB interface in the $KaiBC_{RS}$ complex affects the solvent accessibility into the active site of $KaiC_{RS}$-CI.

(**a**) Size-exclusion chromatography of $KaiB_{RS}$ (blue) shows that it is monomeric in solution in contrast to $KaiB_{SE}$ (gray), which elutes as a tetramer. Molecular-weight standards are shown above the chromatogram. **(b)** Structural comparison of $KaiB_{TE}$ (gray, PDB 5jwq[26]) and $KaiB_{RS}$ (blue) when bound to their corresponding KaiC hexamers. The PISA software package[37] determines that for the $KaiBC_{TE}$ complex the interface between the $KaiB_{TE}$ monomers is 255 $Å^2$, whereas the average interface between $KaiB_{RS}$ monomers is only 45 $Å^2$ in the $KaiBC_{RS}$ complex. **(c)** To understand how $KaiB_{RS}$ binding to $KaiC_{RS}$-CI domain increases the hydrolysis rate, we investigated whether conformational changes modulated substrate access to the active site. The CAVER software[61] was used to calculate tunnels (gray mesh) leading into the active site of the CI domain of $KaiC_{RS}$-S413E/S414E alone (orange) and the $KaiC_{RS}$-S413E/S414E:$KaiB_{RS}$ complex (yellow:blue) with varying probe radii. In both structures, the active site was occupied by ADP:Mg^{2+} (sticks and green sphere, respectively). The crystal structure of bovine F1-ATPase in complex with a transition-state analogue (PDB 1w0j, chain D)[62] was used as a reference to determine the position of the catalytic water molecule in the active site (shown as a red sphere). The calculated tunnels connect bulk solvent to the catalytic water when $KaiB_{RS}$ is bound for probe radii larger than the default value of 0.9 Å, but never in its absence. These results suggest that $KaiB_{RS}$ facilitates the access of water into the active site of $KaiC_{RS}$-CI via long-range conformational changes and thus enhances ATP hydrolysis.

Extended Data Fig. 10 $KaiB_{RS}$ binds preferentially to the post-hydrolysis state of $KaiC_{RS}$ and affects its stability.

(a) Size-exclusion chromatography of 50 µM $KaiC_{RS}$-CI (CI_{RS} domain) in the absence (black line, hexamer) and presence (gray line, monomer) of 50 µM $KaiB_{RS}$ in 1 mM ATP buffer. **(b)** Size-exclusion chromatography of 50 µM $KaiC_{RS}$-Δcoil in the presence of 50 µM $KaiB_{RS}$ (purple). The reference sample (50 µM $KaiC_{RS}$-Δcoil) is a hexamer in solution (cyan) and after the

addition of 50 μM $KaiB_{RS}$ the mixture was incubated at 30 °C for 3.5 h (purple) before running the samples again on a Superdex-200 10/300 GL column at 4 °C. These data show that binding of $KaiB_{RS}$ results in (i) disassembly of the hexameric $KaiC_{RS}$-Δcoil structure into its monomers and (ii) aggregation as detected by the elution in the void volume of the column (v_0). **(c)** Thermal denaturation profiles for $KaiC_{RS}$-S413E/S414E in the presence of 1 mM ADP are shown from dark to light blue for increasing concentrations of $KaiB_{RS}$ (between 0 – 4 μM). The black line represents $KaiB_{RS}$ alone (15 μM), which shows no fluorescence signal as it does not bind to SYPRO Orange due to a lack of a hydrophobic core. The T_m decreases upon the addition of $KaiB_{RS}$, indicating that binding of $KaiB_{RS}$ destabilizes the $KaiC_{RS}$ dodecamer. Likely due to loosening up of interface and the $KaiC_{RS}$ structure, thereby allowing for the formation of a tunnel that connects bulk solvent to the position of the hydrolytic water in the active site (see Extended Data Fig. 9). **(d)** SDS-PAGE analysis showing the control experiment for pull-down assay. The first four lanes after the molecular weight marker are $KaiB_{RS}$-Tag samples (red arrow) and show that $KaiB_{RS}$-Tag binds tightly to the column. The last four lanes are control pull-down assay experiments for $KaiC_{RS}$ (green arrow) and show that $KaiC_{RS}$ alone is unable to bind to the column. The lanes represent the initial sample used in pull-down assay (Before), flow-through after loading sample onto the column (FT), flow-through after washing the column three times with the binding buffer (Wash #3), and sample after elution with imidazole (Elute). (**e**) SDS-PAGE analysis of pull-down assay to measure the complex formation between $KaiB_{RS}$-Tag and wild-type $KaiC_{RS}$, $KaiC_{RS}$-S413E/S414E, or $KaiC_{RS}$-S413A/S414A in the presence of 4 mM ADP or ATP (with an ATP-recycling system). For gel source data, see Supplementary Figure 2. **(f)** Percentage of wild-type $KaiC_{RS}$ bound to $KaiB_{RS}$-Tag protein for different ATP-to-ADP ratios (4 mM total nucleotide concentration) at 25 °C as measured from pull-down assays. **(g)** Fluorescence anisotropy at 30 °C of unlabeled $KaiB_{RS}$ competitively replacing the fluorophore-labeled $KaiB_{RS}$ ($KaiB_{RS}$-6IAF) from $KaiC_{RS}$-S413E/S414E in the presence of 4 mM ADP (red circles, K_D value of 0.79

± 0.06 μM) and 4 mM ATP with an ATP-recycling system (orange triangles). In the latter experiment no change in anisotropy is observed, indicating that only a small fraction of $KaiB_{RS}$-6IAF is bound under these conditions. The average anisotropy and standard error were calculated from ten replicate measurements. **(h)** The mant-ATPγS or mant-ADP release is shown as bar graphs with observed rates of 4.8 ± 0.2 h^{-1} and 43.6 ± 3.0 h^{-1} for mant-ATPγS and mant-ADP releasing from $KaiC_{RS}$-S413E/S414E, respectively, and 21.0 ± 3.0 h^{-1} and 65 ± 15 h^{-1} for mant-ATPγS and mant-ADP releasing from $KaiC_{RS}$-S413A/S414A, respectively. The result shows that CII domain of $KaiC_{RS}$ prefers binding of ATP over ADP. Experiments in panels (f) and (h) were performed in triplicate and data are presented as mean values ± s.d.

Extended Data Table 1 X-ray crystallography data collection and refinement statistics
Extended Data Table 2 Cryo-EM data collection, refinement and validation statistics

Supplementary information

Supplementary Information

Supplementary Figs. 1 and 2 and Supplementary Tables 1 and 2.

Reporting Summary

Supplementary Video 1

Visualization of conformational changes in the CII domain of $KaiC_{RS}$ after phosphorylation through the long-range allosteric network. The trajectory was made by linear interpolation between unphosphorylated $KaiC_{RS}$ and the phosphomimetic variant ($KaiC_{RS}$-S413E/S414E) using the morph tool in UCSF Chimera 1.15 (ref. [64]); the movie was rendered in PyMOL (v.2.6.0)[60].

Supplementary Data 1

Initial multiple-sequence alignment and phylogenetic tree for KaiC.

Supplementary Data 2

Final multiple-sequence alignment and phylogenetic tree including only species that contain KaiB and KaiC.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Pitsawong, W., Pádua, R.A.P., Grant, T. *et al.* From primordial clocks to circadian oscillators. *Nature* **616**, 183–189 (2023).
https://doi.org/10.1038/s41586-023-05836-9

- Received: 29 May 2022
- Accepted: 13 February 2023
- Published: 22 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05836-9

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

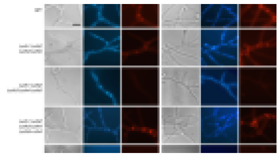
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

The mechanism of the simplest biological 24-hour clock

Nature Research Briefing 22 Mar 2023

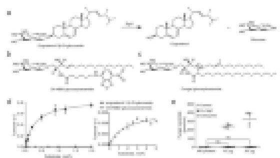
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05836-9



Mechanisms of redundancy and specificity of the Aspergillus fumigatus Crh transglycosylases

10 April 2019

Wenxia Fang, Ana Belén Sanz, … Daan M. F. van Aalten



Structure and inhibition of Cryptococcus neoformans sterylglucosidase to develop antifungal agents

07 October 2021

Nivea Pereira de Sa, Adam Taouil, … Michael V. Airola



Novel Type III Polyketide Synthases Biosynthesize Methylated Polyketides in Mycobacterium marinum

25 April 2018

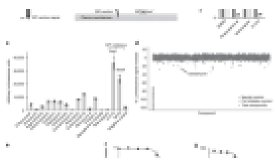
Amreesh Parvez, Samir Giri, … Priti Saxena



[Disrupting coupling within mycobacterial F-ATP synthases subunit ε causes dysregulated energy production and cell wall biosynthesis]

14 November 2019

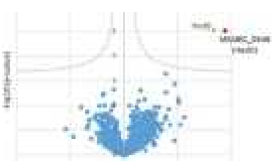
Wuan-Geok Saw, Mu-Lu Wu, … Gerhard Grüber



[Jawsamycin exhibits in vivo antifungal properties by inhibiting Spt14/Gpi3-mediated biosynthesis of glycosylphosphatidylinositol]

07 July 2020

Yue Fu, David Estoppey, … Dominic Hoepfner



[HadD, a novel fatty acid synthase type II protein, is essential for alpha- and epoxy-mycolic acid biosynthesis and mycobacterial fitness]

16 April 2018

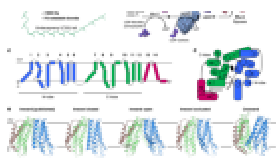
Cyril Lefebvre, Richard Boulon, … Annaïk Quémard



[Identification of Two Mannosyltransferases Contributing to Biosynthesis of the Fungal-type Galactomannan α-Core-Mannan Structure in Aspergillus fumigatus]

16 November 2018

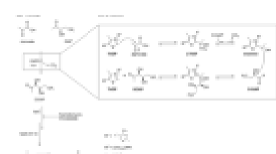
Takuya Onoue, Yutaka Tanaka, … Takuji Oka



[Visualizing conformation transitions of the Lipid II flippase MurJ]

15 April 2019

Alvin C. Y. Kuk, Aili Hao, … Seok-Yong Lee



[First crystal structures of 1-deoxy-d-xylulose 5-phosphate synthase (DXPS) from Mycobacterium tuberculosis indicate a distinct mechanism of intermediate stabilization]

04 May 2022

Robin M. Gierse, Rick Oerlemans, … Matthew R. Groves

- Article

- Published: 22 March 2023

Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1

- Xinlin Hu,
- Ping Yang,
- Changdong Chai,
- Jia Liu,
- Huanhuan Sun,
- Yanan Wu,
- Mingjie Zhang,
- Min Zhang ,
- Xiaotian Liu &
- …
- Hongjun Yu

Show authors

Nature **volume 616**, pages 190–198 (2023)

- 7241 Accesses
- 50 Altmetric
- Metrics details

Abstract

The membrane-integrated synthase FKS is involved in the biosynthesis of β-1,3-glucan, the core component of the fungal cell wall[1,2]. FKS is the target of widely prescribed antifungal drugs, including echinocandin and ibrexafungerp[3,4]. Unfortunately, the mechanism of action of FKS remains

enigmatic and this has hampered development of more effective medicines targeting the enzyme. Here we present the cryo-electron microscopy structures of *Saccharomyces cerevisiae* FKS1 and the echinocandin-resistant mutant FKS1(S643P). These structures reveal the active site of the enzyme at the membrane–cytoplasm interface and a glucan translocation path spanning the membrane bilayer. Multiple bound lipids and notable membrane distortions are observed in the FKS1 structures, suggesting active FKS1–membrane interactions. Echinocandin-resistant mutations are clustered at a region near TM5–6 and TM8 of FKS1. The structure of FKS1(S643P) reveals altered lipid arrangements in this region, suggesting a drug-resistant mechanism of the mutant enzyme. The structures, the catalytic mechanism and the molecular insights into drug-resistant mutations of FKS1 revealed in this study advance the mechanistic understanding of fungal β-1,3-glucan biosynthesis and establish a foundation for developing new antifungal drugs by targeting FKS.

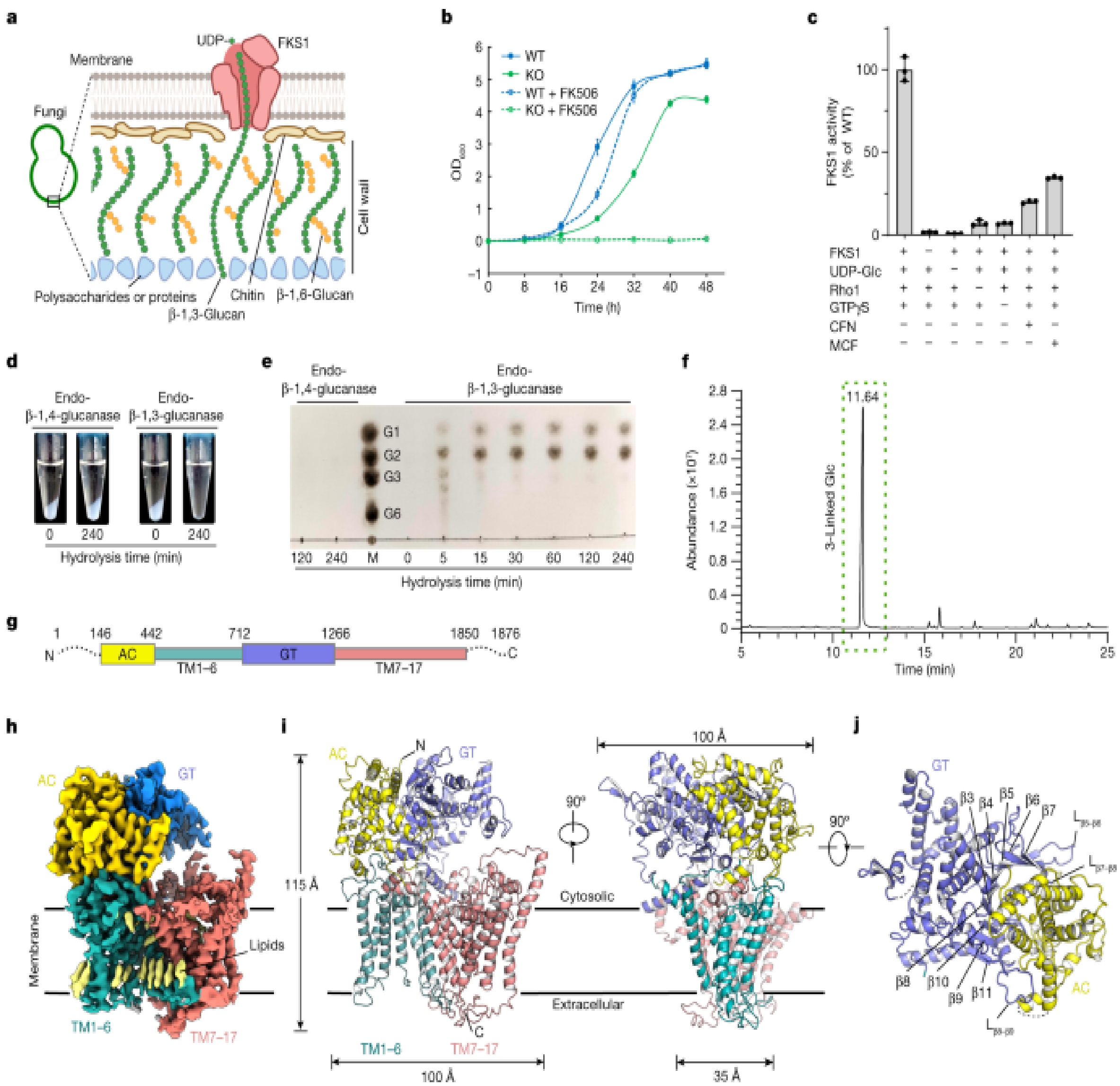
Main

Invasive fungal infections cause over 1.5 million death annually, posing a serious threat to public health[5]. Such infections are of substantial concern (for example, causing high mortality) for immunocompromised populations as well as patients with COVID-19 (refs. [5,6]) (https://www.cdc.gov/fungal/covid-fungal.html). Limited classes of antifungal drugs and emerging drug-resistant strains such as the recent outbreak of multidrug-resistant *Candida auris* have caused serious challenges in treating patients who are infected[7,8]. Therefore, there are pressing needs to develop new antifungal agents[8].

β-1,3-Glucan is a fundamental component of the fungal cell wall[2], thus targeting its biosynthesis is an important strategy for developing broad-spectrum antifungal drugs[3,9]. β-1,3-Glucan of the fungal cell wall is synthesized by membrane-integrated synthase FKS[10,11], which is regulated by GTPase Rho1 (refs. [12,13]). FKS transfers glucose from donor UDP-glucose to the growing chain of glucan via β-1,3-glycosidic linkages and translocates the polymerized β-1,3-glucan into the extracellular space (Fig. 1a). FKS orthologues have been identified in all analysed fungi and are

essential for the viability of many prominent fungal pathogens. In *Candida glabrata* and *S. cerevisiae*, simultaneous disruption of *FKS1* and *FKS2* showed lethal phenotype[14,15]; *FKS1* of *Candida albicans* and *Cryptococcus neoformans* are essential for viability[16,17]; the mould *Aspergillus fumigatus* with an *FKS1* deletion suffers from severe growth defects and cell lysis[18]. Currently, there are two well-known antifungal drugs on the market that target FKS: echinocandin, the first-line antifungal drug widely used in clinical practice[19], and ibrexafungerp, a newly approved oral fungicidal drug[4]. More novel agents targeting FKS function (for example, rezafungin) are entering into late-stage clinical trials[4]. Unfortunately, FKS mutations are closely linked to echinocandin resistance and therapeutic failure, an emerging concern in invasive fungal infections[20,21]. Numerous clinically identified echinocandin-resistant mutations are concentrated at three conserved regions of FKS[20,22]. Mutations in a single *FKS* allele are sufficient to render fungal strains resistant to echinocandin[23].

Fig. 1: Functional and structural characterization of *S. cerevisiae* FKS1.



a, Model of the fungal cell wall. FKS1 uses UDP-Glc to synthesize β-1,3-glucan. The graphic in panel **a** was created using BioRender (https://biorender.com). **b**, In vivo assay of FKS1 by examining the growth of the WT strain and the *FKS1*-KO (KO) strain with or without the FKS2 inhibitor FK506. Data are mean ± s.e.m.; $n = 3$ independent experiments. **c**, In vitro activity of CHAPS-purified FKS1 with different effectors: UDP-Glc, Rho1, GTPγS and antifungal drugs (caspofungin (CFN) and micafungin (MCF)). Activity was measured by monitoring the UDP generated. Data are mean ± s.e.m.; $n = 3$ independent experiments. **d**,**e**, Enzymatic digestion of the water-insoluble polymer synthesized by GDN-purified FKS1(S643P). Specific endo-1,3-β-glucanase and endo-1,4-β-glucanase were used. These experiments were repeated three times with similar results. A representative

image of the polymer digestion (**d**) and hydrolysis samples in **d** withdrawn at indicated times and subject to thin-layer chromatography analysis (**e**) are shown. Glucose (G1), laminaribiose (G2), laminaritriose (G3) and laminarihexaose (G6) were used as the standards (lane M). **f**, Glycosyl linkage (methylation) analysis of the polymer synthesized by GDN-purified FKS1(S643P). A gas chromatogram profile of the partially methylated alditol acetates derived from the synthesized polymers is shown. The dominant peak (outlined by a green dashed box) represents the 1,3-Glc*p* linkage, confirmed by mass spectrometry (Extended Data Fig. 2d). **g**, Schematic of the domain organization of FKS1. **h**, Cryo-EM map of FKS1, viewed parallel to the membrane. The map is segmented into four parts as coloured in **g** with lipid-like densities in light yellow. **i**, Two orthogonal views of the FKS1 structure in cartoon display. Four parts of FKS1 are coloured as in **g**. **j**, Intracellular view of the cytoplasmic region with the TM domain hidden. The central β-sheet of the FKS1 GT domain is composed of β3–β11 as labelled. The dashed lines indicate disordered regions.

The lack of a FKS structure has hindered our understanding of the catalytic mechanism of FKS and hampered developments of new antifungal drugs. FKS is an approximately 200-kDa membrane-embedded protein with low-sequence homologies to other characterized glycosyltransferases (GTs). FKS belongs to the GT48 family with no structures available for any of its members. Here we report the cryo-electron microscopy (cryo-EM) structures of *S. cerevisiae* FKS1 and its drug-resistant mutant FKS1(S643P). Further structure-guided functional characterizations enable us to elucidate the molecular basis of β-1,3-glucan biosynthesis by FKS1 and to rationalize drug-resistant mutations of the enzyme.

Functional characterization of FKS1

For in vivo functional studies, we combined genetic deletion of *FKS1* with pharmacological manipulation of FKS2 to circumvent the lethal phenotype caused by simultaneous disruption of both FKS1 and FKS2 in *S. cerevisiae*[15]. The wild-type (WT) strain showed modestly slowed growth with the FKS2-specific inhibitor FK506 (refs. [15,24]) (Fig. 1b). By contrast, the *FKS1*-knockout (KO) strain exhibited significantly reduced growth without FK506 and a near total growth inhibition with FK506.

Next, we immunopurified the endogenously expressed *S. cerevisiae* FKS1 in the detergent CHAPS, which is frequently used for the FKS1 activity assay[13,25] (see Methods). Traditionally, FKS activity is measured by a radioactive method on crude membrane or product-entrapped FKS[13,25,26], and here we found that monitoring the UDP generated can make a practical radiolabel-free system for the FKS1 activity assay. FKS1 alone did not exhibit apparent activity. The addition of purified Rho1 in the GTPγS-loaded form dramatically boosted the activity of FKS1 (Fig. 1c), confirming the essential regulatory role of Rho1 on FKS1 (refs. [12,13]). The specific activity of WT FKS1 was determined as approximately 824 nmol min^{-1} mg^{-1}, which is comparable with previous studies[10,25,27]. More importantly, the addition of caspofungin or micafungin, two widely used antifungal drugs that can inhibit β-1,3-glucan synthesis[19], significantly reduced FKS1 activity.

To further confirm the role of FKS1 in β-1,3-glucan synthesis, we analysed the product formed by purified FKS1. For this purpose, we tested WT FKS1 and the variant FKS1(S643P), which is one of the most frequently observed echinocandin-resistant mutations[28]. We hypothesized that the latter variant may alter synthase behaviour. When purified to homogeneity in the detergent GDN (Extended Data Fig. 1a–d and Supplementary Fig. 1), both WT FKS1 and FKS1(S643P) exhibited activity, which can be depleted by immunodepletion (Extended Data Fig. 1h–j). We found that the GDN-purified FKS1 can be activated by caspofungin. Caspofungin can stimulate the activity of FKS1(S643P) (Extended Data Fig. 1h–j). This is further confirmed by scale-up product synthesis in which only FKS1(S643P) with caspofungin produced a significant amount of water-insoluble polymer (Extended Data Fig. 1k). The fact that the enzyme purified in GDN and in CHAPS behaves very differently (Extended Data Fig. 1i versus Fig. 1c) suggests that FKS1 function may be sensitive to membrane environments. The product synthesis by FKS1(S643P) exhibits donor specificity towards the preferred ligand UDP-Glc over UDP-GlcNAc (Extended Data Fig. 1l). Moreover, EDTA can support the catalytic activity of the enzyme and Mg^{2+} significantly inhibited the activity (Extended Data Fig. 1m), consistent with previous studies[25,29,30]. Introduction of a K1261A mutation, an enzyme inactivation mutation discovered from our structural analysis (see the section ‘The active site and catalytic mechanism’), to FKS1(S643P) completely inactivated the product synthesis activity of the hyperactive enzyme

(Extended Data Fig. 1h–k). The above assays illustrate the specific enzymatic activity of FKS1.

To confirm the formation of the β-1,3-glycosidic linkage, we performed a hydrolysis assay on the synthesized polymer by FKS1(S643P). As a result, endo-1,3-β-glucanase can completely hydrolyse the polymer, whereas endo-1,4-β-glucanase cannot (Fig. 1d and Extended Data Fig. 2a). Thin-layer chromatography analysis showed that the endo-1,3-β-glucanase-catalysed hydrolysate of our synthesized polymer matches the hydrolysate pattern of curdlan (the β-1,3-glucan standard) (Fig. 1e and Extended Data Fig. 2b,c). In addition, methylation glycosyl linkage analysis showed that 1,3-Glc*p* is the dominant linkage in the synthesized polymer (Fig. 1f and Extended Data Fig. 2d). In summary, the above biochemical and chemical studies together firmly establish that FKS1 is a β-1,3-glucan synthase.

Architecture of FKS1

To better understand the mechanism of FKS1, we determined the cryo-EM structure of GDN-purified FKS1 to an overall resolution of 3.4 Å (Fig. 1g,h, Extended Data Table 1 and Extended Data Fig. 3). The density map allowed the de novo building of an atomic model of FKS1 with most residues resolved (Extended Data Fig. 4). To our knowledge, it represents the first structure of the GT48 family membrane-bound GTs.

The overall structure of FKS1 measures approximately 100 Å × 115 Å × 100 Å (Fig. 1i). It contains a transmembrane (TM) domain with 17 TM helices (TM1–17) and a bulky cytoplasmic entity. The TM domain adopts a wedge-like shape, with its narrower end (approximately 35 Å) at the extracellular side. FKS1 has no apparent extracellular structured domains and several elongated loops connect the TM helices. There are several unstructured regions, detailed in the Methods section ‘Model building and refinement’.

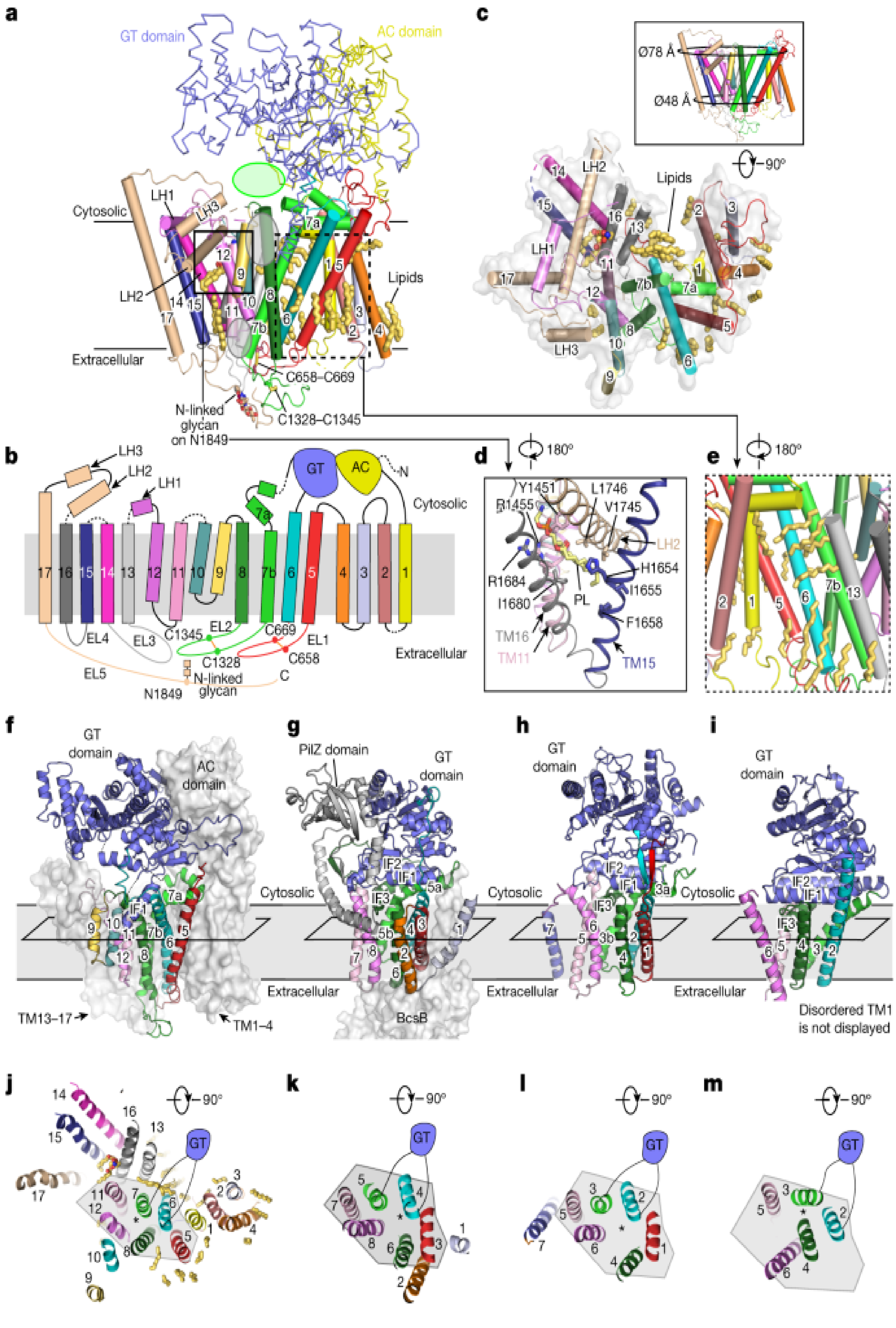
The cytoplasmic part of FKS1 contains an N-terminal domain and a middle domain, separated by TM1–6 in the primary sequence (Fig. 1g). The middle domain (residues 712–1266) adopts a GT-A fold characteristic of GTs: a continuous central β-sheet of nine strands (β3–β11) surrounded by multiple α-helices[31] (Fig. 1j and Extended Data Fig. 3h). It is therefore named as the

GT domain. The N-terminal domain (residues 146–442) contains 14 α-helices (helix 6–7 is not modelled due to lack of densities) (Extended Data Fig. 3i). A search with the Dali server failed to retrieve any structures similar to this domain. Thus, we named it as the accessory (AC) domain. The AC domain interacts extensively with the GT domain, burying a total interface surface area of 2,504 Å^2. Three loops from the GT domain (loop connecting β5 and β6 ($L_{\beta5–\beta6}$), $L_{\beta7–\beta8}$ and $L_{\beta8–\beta9}$) function like plier jaws to clamp the AC domain (Fig. 1j). The highly conserved R319 from the AC domain interacts with N1087 from the GT domain in the domain interface (Extended Data Fig. 3j). The R319A substitution has been previously demonstrated to significantly reduce FKS1 function[24], indicating that the two cytoplasmic domains function together as a compact structural unit.

TM domain of FKS1

Of the total 17 TM helices in FKS1, TM1–8 provide the bulk of the interacting interface with the cytoplasmic entity (Fig. 2a,b and Extended Data Fig. 3i). TM1–4 form a tightly packed helix bundle, contacting TM5–7 on one side (Fig. 2c). TM5–17 are all tilted like open flower petals when viewing from the cytosolic side (Fig. 2a,c). Between the tilted TM5–17 and the GT domain is a large solvent-exposed chamber, which is part of the active site of the enzyme (see below). The TM domain features two membrane-embedded pockets (Fig. 2a). The pocket on the extracellular side is formed because membrane-exposed TM9–10 and TM12 are significantly shorter than typical TM helices. The other pocket on the cytosolic side is formed due to the tilting of TM helices. These two pockets together probably form a path for glucan translocation (see below). Moreover, five long extracellular loops (EL1–5) intertwine to stabilize the close extracellular ends of TM5–17. Two conserved disulfide bonds (C658–C669 on EL1 and C1328–C1345 on EL2) and one potential N-linked glycan on the conserved N1849 are found in these loops (Fig. 2a,b). The tilting of TM5–17 results in large distances between neighbouring helices at the cytosolic end (Fig. 2c). The resulting widespread spaces created are filled in by three lateral helices (LH1–3) parallel to the membrane bilayer: LH1–2 fill the large gap created by two three-helix sheets (TM11, TM12 and TM16; and TM14, TM15 and TM17), whereas LH3 pads the ends of LH1–2.

Fig. 2: The FKS1 TM domain and structural comparison with the cellulose synthases BcsA, CesA and the hyaluronan synthase HAS.



a
GT domain
AC domain
Cytosolic
LH1
LH3
LH2
Lipids
Extracellular
C658–C669
N-linked glycan on N1849
C1328–C1345
b
LH3
LH2
LH1
GT
AC
N
Cytosolic
Extracellular
EL1
EL2
EL3
EL4
EL5
C1345
C669
C1328
C658
N1849
N-linked glycan
C
c
Ø78 Å
Ø48 Å
90°
Lipids
d
180°
Y1451
L1746
R1455
V1745
LH2
H1654
R1684
I1655
PL
I1680
F1658
TM16
TM11
TM15
e
180°
f
GT domain
AC domain
Cytosolic
Extracellular
TM13–17
TM1–4
g
PilZ domain
GT domain
Cytosolic
Extracellular
BcsB
h
GT domain
Cytosolic
Extracellular
i
GT domain
Cytosolic
Extracellular
Disordered TM1 is not displayed
j
90°
GT
k
90°
GT
l
90°
GT
m
90°
GT

a, Front view of FKS1. Seventeen TM helices (denoted by numbers) are in cartoon representation; the cytoplasmic AC and GT domains are in yellow and blue ribbons, respectively. The conserved disulfide bonds at EL1–2 and the N-linked glycan on conserved N1849 of EL5 are labelled. Ordered lipids are in yellow sticks. The three ovals mark the positions of three pockets implicated in catalysis (one green pocket) and glucan translocation (two grey pockets). **b**, Topological diagram of the FKS1 TM domain. The dashed lines depict the flexible elements invisible in the map. **c**, Cytosolic view of the FKS1 TM helices. The inset shows the same view as in **a** with the cytosolic region hidden. The two circles with approximate diameters highlight the tilting of TM5–17, opening towards the cytoplasm. **d**, Residues (shown as sticks) interacting with the internal phospholipid (PL) are outlined by the solid box in **a**. **e**, Back membrane view of the lipid-enriched region outlined by the dashed box in **a**. **f**–**i**, Comparison of the putative shared folds between FKS1 (GT48 family) (**f**) and three membrane-bound synthases of the GT2 family: the bacterial cellulose synthase BcsA (Protein Data Bank (PDB) ID: 4hg6) (**g**), the plant cellulose synthase CesA (PDB ID: 6wlb) (**h**) and the hyaluronan synthase HAS (PDB ID: 7sp7) (**i**). The superimposed structures are shown separately in cartoon representation. Their corresponding elements are in the same colour scheme, whereas elements out of the shared fold are in grey surface or cartoon representation as labelled. In **f**, the dashed lines indicate disordered regions. **j**–**m**, Comparison of the TM arrangements in FKS1 (**j**), BcsA (**k**), CesA (**l**) and HAS (**m**). TM segments in the grey shadow belong to the shared fold. The TM helices are denoted by numbers. The asterisk marks the proposed glucan translocation path in FKS1, BcsA, CesA and HAS.

The resolution of the FKS1 map enabled us to identify multiple orderly bound lipids (Fig. 2a and Extended Data Fig. 4g,h). A phospholipid at the cytosolic membrane leaflet is contacted by TM11, TM15–16 and LH2 (Fig. 2d). Its headgroup is poised to contact the charged residues R1455 (TM11) and R1684 (TM16). Its two lipid alkyl tails pack with surrounding hydrophobic residues: H1654, I1655 and F1658 (TM15); I1680 (TM16); and V1745 and L1746 (LH2). Besides LH1–2, this phospholipid also appears to contribute to the stabilization of the tilted arrangement of the TM helices (Fig. 2a). Other lipid-like densities were modelled as 24 lipid alkyl chains, predominantly located around a core composed of TM5–13 (Fig. 2c). Of these, seven lipids wrap around TM5–6 (Fig. 2c), and 12 form two layers

filling in the concave surface created by the tilted TM helices (Fig. 2e). These well-defined lipid molecules are probably part of the integral structure of FKS1.

A 'cellulose-synthase-like fold' in FKS1

Although the architecture of FKS1 is radically different from the membrane-bound GTs with known structures, we noticed that a part of the FKS1 structure (residues 615–1510) resembles the overall topologies of two cellulose synthases of the GT2 family: bacteria BcsA and its plant homologue CesA[32,33] (Fig. 2f–h). Despite their low-sequence identities (less than 10%), FKS1 and BcsA can be superimposed with a root-mean-square-deviation value of 4.3 Å over 308 aligned residues (Extended Data Fig. 5a). A DALI homologous structure search identified two additional structures resembling this topology: hyaluronan synthase (HAS) and dolichylphosphate mannose transferase (PcManGT). Consistently, HAS has been reported with an overall structural similarity to BcsA[34] (Fig. 2i,m); PcManGT has been proposed to contain a minimal cellulose-synthase-like fold with BcsA[35]. These further implicate that these membrane-bound polysaccharide synthases (FKS1, BcsA, CesA and HAS) may share a core fold, which we refer to as a cellulose-synthase-like fold. This fold shares several features: it is composed of six TM helices, with the GT domain splitting the N-terminal two helices from the C-terminal four helices (Fig. 2j–m); the membrane–cytosol boundary contains several amphipathic interface helices (IFs) (Fig. 2f–i); near this boundary, the TM helix following the GT domain (FKS1 TM7, BcsA TM5, CesA TM3 and HAS TM3) is elongated and has been proposed to line the glucan translocation path[32,33,34].

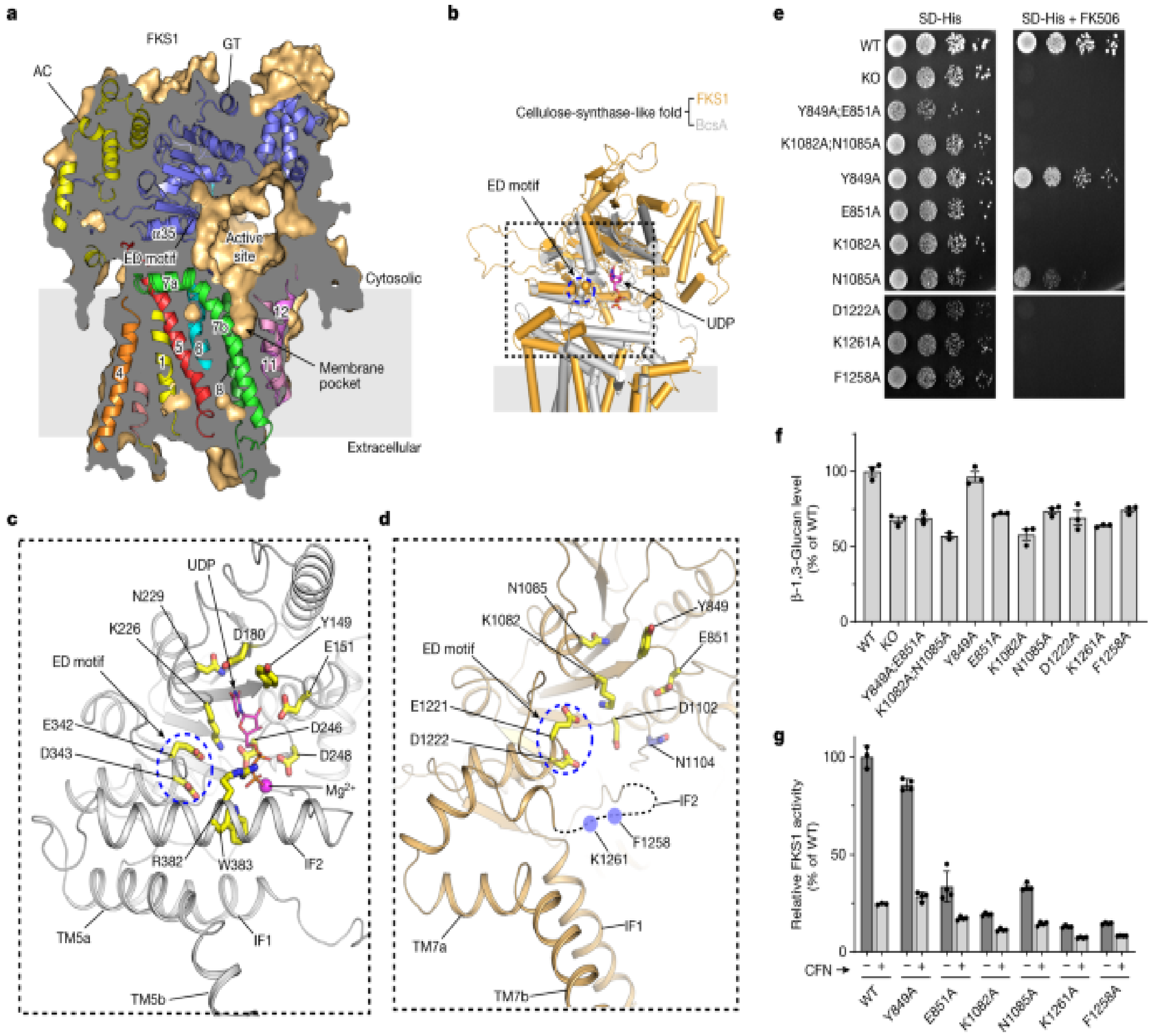
Despite the overall resemblance of this modular fold, considerable FKS1-specific features were revealed (Extended Data Fig. 5). The GT domain gains considerate extensions relative to BcsA (Extended Data Fig. 5a,b), and the topologically corresponding TM helices between FKS1 and BcsA vary in orientation and length (Extended Data Fig. 5c). More notable variations are at the membrane–cytoplasm interface. BcsA, CesA and HAS feature three similar amphipathic IFs (IF1–3)[33,34,36] (Fig. 2g–i). In FKS1, the topological counterpart of IF3 becomes TM helices (TM9–10) instead

(Extended Data Fig. 5d,e). Although IF1–2 are preserved in FKS1, they show distinct states. IF1 (that is, helix α33 of the GT domain) of FKS1 rotates almost 90° and is sandwiched between TM6 and TM8 (Extended Data Fig. 5d). IF2 of BcsA corresponds to residues 1247–1266 of FKS1, which are disordered in the map, although predicted as a helix (Extended Data Fig. 5c). This putative FKS1 IF2 lacks the 'QxxRW' motif conserved among BcsA, CesA and HAS[33,34,36] but could still contribute to catalysis, as we discuss later.

The active site and catalytic mechanism

Within the cellulose-synthase-like fold, FKS1 contains a large solvent-exposed chamber at the membrane–cytoplasm interface (Fig. 3a). Alignment of this fold with that of BcsA reveals that the chamber in FKS1 overlaps well with the active site of UDP-bound BcsA, strongly indicating the chamber as the active site of FKS1 (Fig. 3b). A DALI search retrieved homologous structures of the FKS1 GT domain (residues 718–1239): PcManGT, BcsA, GalNAc-T7, TarP, CesA and HAS. We also analysed the AlphaFold model of curdlan synthase because it synthesizes the same polymer as FKS1 and shows high homology (26% identities) with BcsA[37]. Detailed structural comparisons of these targets revealed that the active site of FKS1 overlaps better with membrane-bound enzymes such as BcsA, CesA, curdlan synthase, HAS and PcManGT (Extended Data Fig. 6). Despite the different stereochemistry of their products, the high structural similarity of their active sites indicates similar mechanisms in substrate binding and catalysis of these synthases, consistent with previous studies of BcsA, CesA and HAS[33,34,36]. On the basis of these insights, we overlaid the structures of FKS1 and BcsA to reveal the shared key functionally important residues in their active sites (Fig. 3c,d). For example, two pairs of residues (Y849 and E851, and K1082 and N1085 in FKS1; Y149 and E151, and K226 and N229 in BcsA) and the 'ED motif' (E1221 and D1222 in FKS1; and E342 and D343 in BcsA) in the two enzymes can be aligned well. These residues in BcsA are known for donor binding (Y149 and E151, and K226 and N229 in BcsA) and for positioning the general catalytic base aspartate (ED motif)[36].

Fig. 3: The active site of FKS1.



a, A cutaway surface representation of FKS1 showing the active site at the membrane–cytoplasm interface. It connects a membrane pocket, which has the potential for glucan translocation. Specific structural elements are shown in cartoon representation, coloured as in Fig. 2a. The TM helices are denoted by numbers. The highly conserved ED motif (E1221 and D1222) located on helix α35 is facing the active site as marked. **b**, Superimposed GT domains of the cellulose-synthase-like folds shared between UDP-bound BcsA (PDB ID: 4P00; in grey with UDP in the magenta stick) and FKS1 (in orange). **c**,**d**, Close-up view of the active sites between BcsA (**c**) and FKS1 (**d**), as marked by the dashed box in **b**. Residues involved in substrate binding or catalysis are indicated as sticks. In **c**, UDP is shown as the magenta stick. In **d**, FKS1 residues conserved with BcsA are highlighted as yellow sticks. **e**, In vivo functional assay of active site mutations in FKS1. The growth of the

indicated strains for 72 h was assayed without (left rank) and with (right rank) the FKS2-specific inhibitor FK506. The test was repeated three times with similar results. **f**, Levels of cell wall glucan in strains carrying the indicated FKS1 mutations. **g**, In vitro FKS1 activities of WT and active site mutants purified with CHAPS, with or without the FKS1 inhibitor CFN. Activities were normalized to the amount of protein. For **f**,**g**, data are mean ± s.e.m.; $n = 3$ independent experiments.

Next, we validated the functional roles of these residues using three different assays. First, we introduced their mutations into the *FKS1* chromosomal locus and assayed the growth phenotype of each mutant strain in the presence of FK506 (Fig. 3e). The double mutations of Y849A;E851A and K1082A;N1085A and the single mutations of E851A, K1082A and D1222A all prevented strain growth to the extent of the *FKS1*-KO strain (Fig. 3e). The N1085A mutation slowed growth and the Y849A mutation did not have an obvious effect on growth. Second, we measured the levels of cell wall glucan in these strains when growing without FK506. We found that all mutations, including the *FKS1*-KO strain but except for the Y849A substitution, decreased the overall levels of glucan compared with the WT strain (Fig. 3f), matching well with their growth phenotypes (Fig. 3e). Last, we purified every single FKS1 mutant (except for FKS1(D1222A)) and compared their catalytic activity in vitro (Fig. 3g and Extended Data Fig. 1g). Consistent with the cell-based phenotypes described above, the E851A, K1082A and N1085A mutations significantly reduced the enzyme activity, whereas the Y849A mutation only mildly affected the activity. Together, the above in vivo and in vitro analyses firmly establish the location of the FKS1 active site as well as its key composing residues.

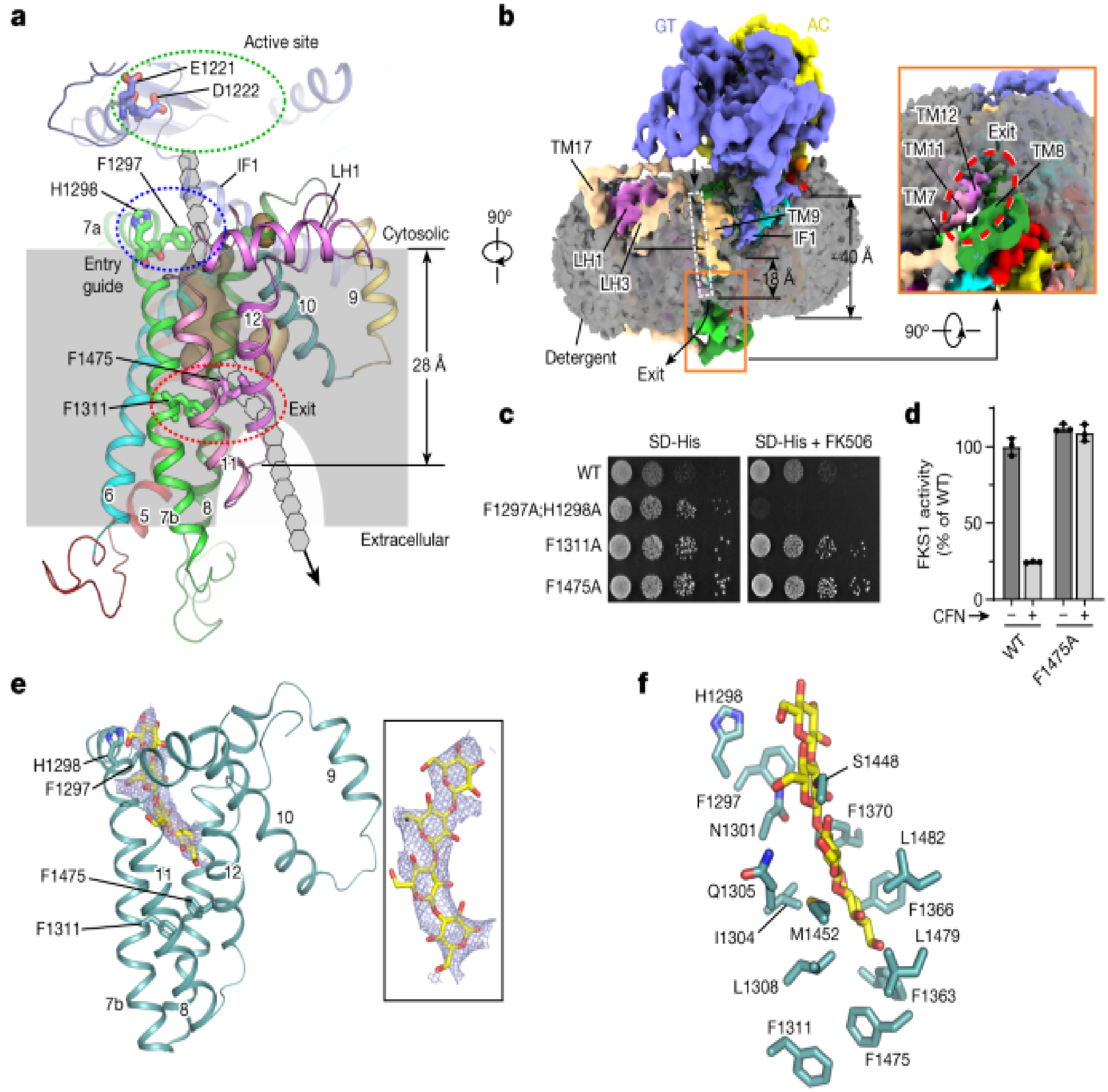
Further structural analysis allowed us to identify additional residues critical for FKS1 catalysis. R382 and W383 in the 'QxxRW' motif in IF2 of BcsA is known to be important for substrate binding and catalysis[36]. Although this motif is missing in FKS1, K1261 and F1258 are located at the corresponding positions in FKS1, as R382 and W383 do in BcsA (Fig. 3c,d and Extended Data Fig. 5e). In multiple membrane-bound synthases containing the cellulose-synthase-like fold (BcsA, CesA and HAS), the positively charged residue corresponding to BcsA R382 is involved in donor binding, whereas the aromatic residue corresponding to BcsA W383 functions to stabilize the acceptor[33,34,36] (Extended Data Fig. 6). Indeed, both F1258 and K1261 are

essential for FKS1 function, as demonstrated by the in vivo and in vitro analyses of their single alanine mutations (Fig. 3e–g). The above analyses further suggest that the core catalytic mechanisms between FKS1 and BcsA are conserved.

The putative glucan translocation path

FKS1 and BcsA are both processive polysaccharide synthases and their polymerized product, glucans, need to be translocated across membranes[1,38]. In accordance with this, right below the active site, FKS1 has a pocket formed within the membrane (Fig. 3a). It is lined by TM5–8, TM11–12 and IF1 (Fig. 4a). TM5–8 and TM11–12 of FKS1 are part of the cellulose-synthase-like fold core and they correspond to TM3–8 forming the cellulose translocation path in BcsA[36] (Fig. 2j,k). At the extracellular side opposite to this pocket, there exists another solvent exposed pocket with much lower detergent density that is directly visible from the EM map (Fig. 4b). The second pocket is formed because the membrane-exposed TM9–10 and TM12 are short. The two juxtaposed hydrophilic pockets with the membrane plane lead to notable membrane thinning (Fig. 4a,b), which could greatly reduce the energy barrier for glucan translocation across the membranes, as has been proposed for several membrane translocases[39]. Indeed, mutations of residues along this path in FKS1 altered FKS1 function (Fig. 4c,d). For example, two conserved residues (F1297 and H1298) are located between the active site and the cytosolic entrance. Their alanine substitutions impaired FKS1 function, suggesting that they may be involved in glucan guiding. By contrast, when substituting two conserved bulky aromatic residues sealing the extracellular exit (F1311 and F1475) with alanine, the mutant strains grew faster than the WT strain. We were able to purify FKS1(F1475A) and it exhibited increased in vitro catalytic activity and reduced sensitivity to the FKS1 inhibitor caspofungin (Fig. 4d), implying the enlargement of F1475A in the glucan exit.

Fig. 4: Putative glucan translocation path.



a, Putative glucan translocation path through two membrane pockets as illustrated by the hexagon array. The first pocket (in the internal shaded surface view), lined by TM5–12, is located under the active site (green dashed oval). Conserved residues implicated in guiding the entrance (H1298 and F1297) and sealing the exit (F1311 and F1475) are labelled. At the extracellular side, a second pocket is formed because membrane-exposed TM9–10 and TM12 are too short (less than 28 Å) to span the whole membrane. **b**, Unsharpened EM density map shows the detergent depression close to the glucan translocation path. This path is marked by the white dashed box, with the arrows indicating the putative entrance and exit. The detergent density is in grey. The estimated membrane thickness is labelled.

The inset shows the extracellular view of the region boxed with the orange line. **c**, In vivo functional assay of FKS1 with selected mutations around the glucan translocation path. The growth of the indicated strains for 48 h were assayed without (left) and with (right) the FKS2-specific inhibitor FK506. **d**, In vitro FKS1 activities of the WT and the F1475A mutant purified with CHAPS, with or without the FKS1 inhibitor CFN. Activities were normalized to the amount of protein. Data are mean ± s.e.m.; $n = 3$ independent experiments. **e**, Structure of FKS1(S643P) shows an elongated density (blue mesh) in the first membrane pocket (as shown in **a**) for product translocation. This structure was determined in the presence of UDP-Glc to synthesize product, which was boosted by CFN. The TM helices are denoted by numbers. The inset shows the zoomed-in view of a modelled glucan chain of 4-glucose (yellow stick) superimposed with its corresponding density (blue mesh). **f**, Potential residues involved in glucan interaction.

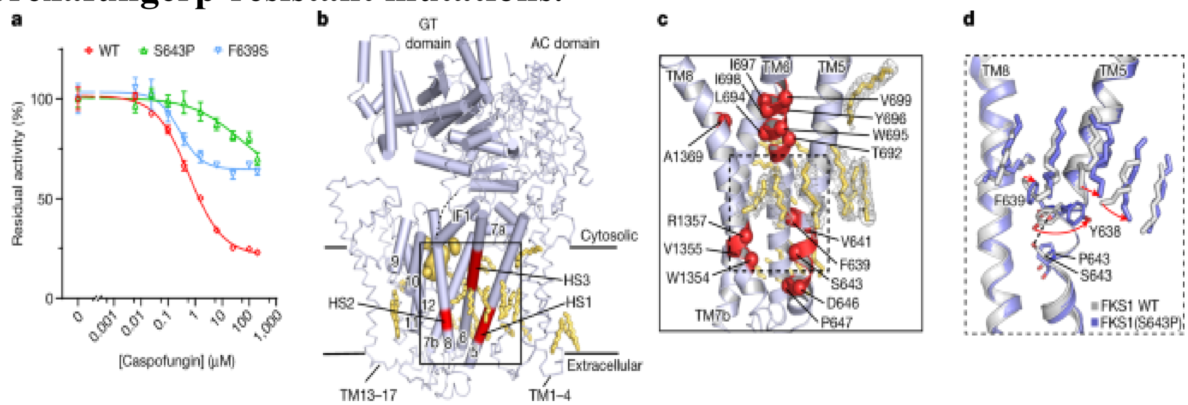
We next determined the structure of the GDN-purified hyperactive mutant FKS1(S643P) when incubated with UDP-Glc and caspofungin, to the resolution of 3.5 Å (Extended Data Fig. 7), hoping to capture the substrate-bound or product-bound state of the enzyme. Although the structure of FKS1(S643P) shows no large conformational differences from that of WT FKS1 (root-mean-square deviation of approximately 0.7 Å), we noticed an elongated density along the proposed product translocation channel in it, presumably corresponding to a bound product (Fig. 4e). We modelled a glucan chain of 4-glucose into this density (Fig. 4e, inset). A series of hydrophilic and hydrophobic residues from FKS1 are lined along the modelled glucan chain for favourable enzyme–product interactions (Fig. 4f). The translocation channel remains closed at the extracellular side, suggesting the existence of a regulated channel opening mechanism, as has been recently proposed for HAS[34].

Implications for echinocandin resistance

We first evaluated our *S. cerevisiae*-based assay system for the analysis of echinocandin resistance. We selected two most frequently observed mutations in drug-resistant *C. albicans* (*Ca*FKS1 F641S and S645P)[28] and introduced them into the corresponding sites of *S. cerevisiae* FKS1 (*Sc*FKS1 F639S and S643P). Both mutant strains exhibited highly elevated resistance

to caspofungin (Extended Data Fig. 8a). We further purified these two mutated *Sc*FKS1 in CHAPS and evaluated their inhibition by caspofungin (Fig. 5a). The WT *Sc*FKS1 as control exhibited a typical inhibitory profile, with the IC_{50} determined as 0.55 μM, which is comparable with that determined previously with different methods[23,25]. By contrast, both *Sc*FKS1(F639S) and *Sc*FKS1(S643P) are insensitive to inhibition by caspofungin.

Fig. 5: Structural interpretation of echinocandin-resistant or ibrexafungerp-resistant mutations.



a, CFN inhibition profiles of FKS1 variants purified with CHAPS, including WT FKS1 and two most frequently observed echinocandin-resistant mutations: F639S and S643P. The percentage of residual activity is shown. Data are mean ± s.e.m.; $n = 3$ independent experiments. **b**, Three hotspot regions (red; denoted as HS1–3) with echinocandin-resistant mutations are mapped onto the FKS1 structure. The hotspot regions are localized on a cluster of three neighbouring TM helices: TM5, TM8 and TM6, respectively. The conserved cellulose-synthase-like fold within FKS1 is highlighted in cartoon representation, whereas the remaining part is in ribbon representation. This conserved fold is surrounded by ordered lipids shown as yellow sticks. The dashed line indicates disordered region. **c**, Close-up view of the echinocandin-resistant mutations as outlined in **b**. An ibrexafungerp-resistant mutation site (A1369) outside these hotspot regions is also shown. The Cα atoms of the mutations are denoted as red spheres. The hotspot regions are enriched with ordered lipids, with selected lipid densities shown in a grey mesh. **d**, Conformational changes and lipid arrangements (indicated by red arrows) around the HS1 region due to the introduction of FKS1(S643P), the drug-resistant mutation. This is illustrated

by the overlay of the FKS1 structure (grey) and the FKS1(S643P) structure (blue). The black dashed line indicates a potential polar interaction.

Next, we investigated the possible mechanistic basis of echinocandin-resistance using the FKS1 structure. According to their locations, echinocandin-resistant mutations were categorized into three conserved regions, encompassing residues 639–647 in TM5 (hotspot 1), 1354–1357 in TM8 (hotspot 2) and 690–700 in TM6 (hotspot 3) (residues are numbered as in *Sc*FKS1)[20,22,40] (Supplementary Fig. 2). Although separated in the primary sequence, three mutation hotspots are spatially close to each other (Fig. 5b,c). In *C. glabrata* isolates resistant to ibrexafungerp, new mutation sites outside FKS2 hotspots 1 and 2 were reported (for example, *Cg*FKS2 W715L and A1390D)[41]; these mutation sites correspond to W695 in hotspot 3 of *Sc*FKS1 and A1369 on TM8 adjacent to hotspot 3, respectively (Fig. 5c). This analysis suggests overlapping FKS-binding sites between echinocandin and ibrexafungerp[41]. These drug-resistant hotspot regions appear to be closely related to the membrane environment. All echinocandin-resistant mutations of FKS1 are mapped near the convex hydrophobic surface formed by TM5–6 and TM8 and are specifically surrounded by several ordered lipid molecules (Figs. 2c and 5c). F639 and S643 of the hotspot 1 region are involved in lipid binding; the side chain of F639 directly interacts with three lipid molecules and the S643 side chain appears to fix the main chain of the lipid-interacting residue Y638 (Fig. 5d). Furthermore, we analysed the structure of FKS1(S643P) (Extended Data Fig. 7), focusing on alterations of lipid binding induced by the mutation (Fig. 5d). Several hotspot 1 residues near the enriched lipids show altered conformation; Y638 has its side chain rotated approximately 90°, with the F639 side chain rotating in the same direction. Remarkably, these changes lead to shifts of nearby bound lipids, suggesting that an altered lipid environment resulted from the S643P mutation.

On the basis of our results, we propose two possible drug-resistant mechanisms for FKS1. First, as the echinocandin-type drugs are lipopeptides with differently configured lipid tails[19] (Extended Data Fig. 8b,c), the clustering pattern of the echinocandin-resistant mutations suggests a possible binding site for these drugs, and the bindings may involve their lipid tails (Fig. 5b,c). The drug resistance may hence be attributed to alterations of the

drug-binding site that resulted from the mutations. Second, considering the enriched ordered lipids near drug-resistant hotspot regions and the lipid movements revealed from the structure of FKS1(S643P) (Fig. 5c,d), it is conceivable that the drug-resistant mutations could alter the response of FKS1 to membrane changes caused by echinocandin. Such a mechanism has been proposed for the antibiotics daptomycin and polymyxins, which are also lipopeptide drugs and membrane-acting[42,43]. Consistently, it has been reported that specific membrane microdomains are associated with the synthesis of β-1,3-glucan[44,45]. In addition, changes in the lipid microenvironment have been demonstrated to correlate with the action of echinocandin on FKS[46,47].

Conclusions

Our work reveals the molecular architecture of the fungal glucan synthase FKS1. The structures of FKS1, together with extensive functional characterizations, advance our mechanistic understanding of fungal cell wall β-1,3-glucan synthesis. Structure-based analysis of echinocandin-resistant mutations and the structure of the S643P mutant suggest a plausible drug-resistant mechanism of the FKS1 mutations. These insights may also be shared by various fungal pathogens as their FKS orthologues are highly conserved (Extended Data Fig. 9 and Supplementary Fig. 2). The FKS1 structures and the catalytic mechanism elucidated in this work may serve as a framework for future studies of this important antifungal drug target, as well as for screening and the development of new antifungal agents.

Methods

Expression and purification of FKS1

A 3× Flag tag was engineered to the C terminus of chromosomal FKS1 on *S. cerevisiae* strain BY4742, using a PCR-based tagging method[48]. A yeast strain with the FKS1(S643P) chromosomal mutation was generated using the homologous recombination method[49]. For cryo-EM analysis and product synthesis, FKS1 and the FKS1(S643P) mutant were purified with detergent GDN as described below. The strains were cultured in YPD medium at 30

°C for 20 h. Cells were collected by centrifugation and lysed by French Press in lysis buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl and 2 mM $MgCl_2$ supplemented with 1 mM phenylmethane sulphonylfluoride (PMSF). The lysate was centrifuged at 15,000*g* for 30 min. The supernatant was subject to ultracentrifugation at 100,000*g* for 1 h. The collected membrane pellet was solubilized in buffer 50 mM Tris-HCl pH 7.4, 500 mM NaCl, 2 mM $MgCl_2$, 10% (v/v) glycerol, 1.5% (w/v) *n*-dodecyl-β-d-maltopyranoside (DDM; Anatrace), 0.15% cholesteryl hemisuccinate Tris salt (CHS; Anatrace) and protease inhibitor (cOmplete protease inhibitor cocktail; Roche) by gentle agitation for 2 h. Supernatant was collected by centrifugation at 15,000*g* for 0.5 h and was applied to Anti-Flag M2 affinity gel (Sigma). The gel was then washed with washing buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM $MgCl_2$ and 0.04% GDN (Anatrace). The target proteins were eluted with washing buffer supplemented with 150 μg ml^{-1} 3× Flag peptide. The eluate was concentrated and further purified using size-exclusion chromatography (Superose 6 10/300 GL column, GE Healthcare) with buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM $MgCl_2$ and 0.04% GDN (Anatrace). The central fractions of monodisperse peak were collected and concentrated for cryo-EM grid preparation and product synthesis.

Expression and purification of Rho1

The vector construction of N-terminal 6× His-SUMO-tagged *S. cerevisiae* Rho1 was transformed into *Escherichia coli* BL21(DE3). The strains were cultured in LB medium supplemented with 100 μg ml^{-1} ampicillin at 37 °C till OD_{600} reached approximately 0.6. Protein expression was induced with 0.5 mM isopropyl-β-d-1-thiogalactopyranoside at 18 °C for 20 h. The collected bacteria were resuspended and lysed by French Press in buffer A containing 50 mM Tris-HCl pH 7.4, 300 mM NaCl and 2 mM $MgCl_2$. After centrifugation, the protein was purified from the supernatant using nickel–nitrilotriacetic acid (Ni–NTA) affinity chromatography. When dialysing into buffer A, 6× His-tagged protease Ulp1 at a 1:200 Ulp1:Rho1 (w/w) ratio was added to remove the 6× His-SUMO tag. The cleaved 6× His-SUMO tag and the protease were removed by another run of Ni–NTA affinity chromatography, and the flow-through was collected, concentrated and

subjected to a gel-filtration chromatography (Superdex 200, GE Healthcare) in buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl and 2 mM $MgCl_2$. The central fractions of the monodisperse peak at 15.6 ml were collected and concentrated to 10 mg ml^{-1}.

Cryo-EM data acquisition

For single-particle cryo-EM analysis, a 3 µl aliquot of the purified FKS1 at a concentration of 5 mg ml^{-1} was applied to glow-discharged Quantifoil carbon grids (R1.2/1.3 Au, 300 mesh). The grids were blotted for 4 s at 100% humidity and flash frozen in liquid ethane using FEI Vitrobot IV. For the UDP-Glc-incubated FKS1(S643P) sample, FKS1(S643P) purified in GDN (7 mg ml^{-1}) was mixed with UDP-Glc (0.5 mM), supplemented with 0.7 mg ml^{-1} Rho1, 10 µM GTPγS, 200 µM caspofungin, 0.1% CHAPS and 0.02% CHS. The mixture was incubated at 16 °C for 2 h before freezing the cryo-EM grids. Cryo-EM data were collected on a FEI Titan Krios electron microscopy operated at 300 kV equipped with a Gatan K3 Summit camera positioned after a GIF quantum energy filter. Automated data acquisition was performed with SerialEM or FEI EPU. Micrographs were recorded under super-resolution counting mode at a nominal magnification of ×130,000, in a physical pixel size of 0.92 Å per pixel. Defocus values ranged from −1.2 µm to −3 µm. A total exposure of 1.3 s was dose-fractionated into 32 frames, resulting in a total accumulated dose of 50 e^- per $Å^2$.

Image processing and 3D reconstruction

The dose-fractionated movies recorded were first aligned and dose-weighted with MotionCor2 (ref. [50]). CTFFIND4 was then used to determine the contrast transfer function parameters for individual micrographs[51]. Low-quality micrographs revealed by manual inspection were excluded from further analysis. Subsequent image-processing steps were performed using RELION-3 (ref. [52]).

For the FKS1 dataset, a set of 2,000 particles were manually selected to generate 2D class templates for reference-based automatic particle picking. The automatic picking yielded in 2,321,905 particles from 11,606

micrographs. Two rounds of reference-free 2D classification were performed to remove particles with poor quality, resulting in a cleaned set of 1,335,014 particles. An ab initio map was generated with RELION and was used as the initial reference model for further 3D classification. After three rounds of 3D classification, a 3D class with high-resolution features (267,574 particles) was selected. Subsequent particle polishing, 3D refinement and post-processing generated a map with an overall resolution of 3.4 Å.

For the FKS1(S643P) dataset, a set of approximately 2,000 particles were manually selected to generate 2D class templates for reference-based automatic particle picking. The automatic picking yielded in 2,222,315 particles from 9,623 micrographs. Two rounds of reference-free 2D classification were performed to remove particles with poor quality, resulting in a cleaned set of 1,620,308 particles. The FKS1 model was used as the initial reference model for further 3D classification. A 3D class with high-resolution features (690,251 particles) was selected. Subsequent particle polishing, 3D refinement and post-processing generated a map with an overall resolution of 3.6 Å. This reconstruction revealed fragment density within the product translocation channel, indicating a mixed state with or without bound product. Therefore, we performed an additional round of focused 3D classification without particle realignment, using a focused mask around TM7–12 and the active site that surround the translocation channel. A 3D class with highest resolution features (176,682 particles) was selected. The selected particles were re-extracted for 3D refinement with full mask and subsequent post-processing. This generated a map with an overall resolution of 3.5 Å, which was used to build the FKS1(S643P) model.

The overall resolutions were estimated based on the gold-standard Fourier shell correlation 0.143 criterion[53]. Local resolution distribution was estimated using ResMap[54].

Model building and refinement

A rough initial model of FKS1 was generated de novo using the map_to_model module in PHENIX[55]. It was further improved by manual adjustments and rebuilding using Coot[56], which was facilitated by the good densities around the TM helices and bulky densities around residues such as

Trp, Tyr, Phe and Arg. The refinement of the FKS1 model against the cryo-EM map in real space was carried out in PHENIX with secondary structure and geometry restraints[55]. In the final FKS1 model, there are several unstructured regions, including N-terminal 145 residues, C-terminal 16 residues, six flexible segments of cytoplasmic domain (residues 244–278, 475–487, 798–805, 897–931, 1159–1167 and 1247–1266) and four loop segments to connect TM helices (residues 1419–1435, 1516–1554, 1627–1637 and 1698–1723). The model of FKS1(S643P) was built based on the model of FKS1. MOLPROBITY was used to assess the final model[57]. The Fourier shell correlation between the model and the map was calculated by PHENIX.mtriage[58]. The homologous structure search was performed using the DALI server[59]. For structural comparison with FKS1, the AlphaFold-modelled structure of curdlan synthase was used[60,61]. The illustrated figures were prepared using PyMOL (Schrödinger, LLC), Chimera and ChimeraX[62,63]. Statistics of the 3D reconstructions and model refinements are shown in Extended Data Table 1.

Spot growth assay

Yeast strains with FKS1 chromosomal mutations and the *FKS1*-KO strain were generated using the homologous recombination method[49]. The mutants were selected on SD-His (Yeast Synthetic Drop-out Medium without Histidine, cat no. S0020, Solarbio) and were confirmed by genomic PCR and DNA sequencing. For the analysis of growth phenotype, the spot growth assay of mutated strains was adapted from a previously described method[24]. The strains were cultured in YPD media at 30 °C at 200 rpm to exponential phase until OD_{600} was 0.8. The cells were diluted to an OD_{600} of 0.1. Then, 5 µl portions of tenfold dilutions of the indicated strains were spotted on SD-His with or without FK506 (1 µg ml^{-1}). After incubation at 30 °C for indicated time, plates were photographed with the 5200CE Image System (TANON). Each assay was repeated three times with similar results.

Growth curve analysis

To profile the cell growth, the WT strain and the *FKS1*-KO strain were grown in YPD media at 30 °C overnight to exponential phase. Then,

cultured cells were inoculated into the fresh YPD medium with an initial OD_{600} of 0.01, which was supplemented with or without 1 μg ml^{-1} FK506. The new cultures were then grown at 30 °C till stationary phase. During the cultivation, the optical densities (OD_{600}) were measured at 8 h interval for 48 h. The experiments were done in triplicate to make the cell growth curves.

FKS1 activity assay

The yeast strains carrying different FKS1 variants with the C-terminal 3× Flag tag were generated, cultured and lysed as described above. The collected membrane was solubilized in buffer containing 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 33% glycerol, 0.5% CHAPS (Anatrace), 0.1% CHS (Anatrace), 4 μM GTPγS and protease inhibitor (cOmplete protease inhibitor cocktail, Roche). FKS1 and its variants were purified using anti-Flag M2 affinity gel (Sigma) and were eluted in buffer containing 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 33% glycerol, 0.2% CHAPS and 0.04% CHS, supplemented with 150 μg ml^{-1} 3× Flag peptide. The proteins purified in detergent CHAPS or GDN were assayed by using the UDP-Glo Glycosyltransferase Assay kit (Promega) to monitor the UDP released. Of FKS variants, 2.5 μl was added to 30 μl reaction mixture in total containing 50 mM Tris-HCl pH 7.4, 33% glycerol, 1 mM EDTA, 6 μg ml^{-1} Rho1, 0.2% CHAPS, 0.04% CHS, 4 μM GTPγS and 20 mM potassium fluoride (KF). The reaction was initiated by adding UDP-Glc to a final concentration of 2.5 mM. The reaction was carried out at 30 °C for 1 h. Luminescence was recorded using a Pherastar FS system (BMG Labtech). The final activity was normalized to the FKS1 amount that was measured by immunoblotting against the 3× Flag tag, using the DYKDDDDK tag monoclonal antibody (1:1,000 dilution; cat no. 66008-3-Ig, clone no. 2B3C4, Proteintech). In the case of inhibitor profiling, serial dilutions of echinocandin drugs were first incubated with FKS1 variants for 10 min at room temperature before adding other components to react.

In vitro synthesis of β-1,3-glucan and aniline blue staining

For in vitro β-1,3-glucan synthesis, the enzyme purified in detergent GDN (FKS1 or the FKS1(S643P) mutant; 0.02 mg ml^{-1}) and donor UDP-Glc (2.5 mM) were mixed in reaction buffer (50 mM Tris-HCl pH 7.4, 33% glycerol, 1 mM EDTA, 6 μg ml^{-1} Rho1, 0.2% CHAPS, 0.04% CHS, 4 μM GTPγS and 20 mM KF), in the presence or absence of 200 μM caspofungin. The reaction volume was set up as 100 μl (for aniline blue staining) or 10 ml (for product enrichment and subsequent enzymatic degradation and glycosyl linkage analysis). The reaction was carried out at 30 °C for indicated time period.

For the staining of the synthesized products by aniline blue, a 100 μl aliquot of reactants or 0.1% (w/v) of *S. cerevisiae* β-glucan (Sigma) was taken and added to equal volume of aniline blue (0.03%; Sigma). The mixture was incubated in the dark for 20 min for complete dye binding. Each sample was then loaded into a capillary and was observed under a fluorescence microscope with an excitation wavelength of 365 nm and an emission wavelength of 433 nm. Each assay was repeated three times with similar results.

Enzymatic degradation analysis of the synthesized polymer

The in vitro synthesized polymer is water-insoluble and was collected by centrifugation. It was washed twice with buffer: 50 mM Tris-HCl pH 7.4, 33% glycerol, 1 mM EDTA, 0.2% CHAPS, 0.04% CHS, 20 mM KF and twice with deionized water. Two enzymes were used for the degradation analysis of the synthesized polymer: endo-β-1,3-glucanase (*Trichoderma* sp.; product code: E-LAMSE, Megazyme) or endo-β-1,4-glucanase (*Aspergillus niger*; product code: E-CELAN, Megazyme). These two enzymes were first dialysed into the buffer (100 mM NaAc pH 4.5). Then, 0.5 U of each enzyme was added to 50 μl 1% (w/v) of synthesized polymer in 200 mM NaAc (pH 4.5). Degradation of standard controls cellulose (C6288, Sigma) or curdlan (C7821, Sigma) was also performed with the same condition[64,65]. The mixtures were incubated at 40 °C, and samples were withdrawn at various time intervals and were boiled for 5 min for reaction termination. The enzymatic degradation products were then analysed by thin-layer chromatography. In brief, a 2 μl aliquot of the withdrawn samples was spotted on a silica gel plate (Merck Silica gel 60

F254) and developed in a solvent system containing *n*-butanol:acetic acid:water (2:1:1 v/v/v). The plates were visualized by immersing in methanol:sulfuric acid (95:5 v/v) and subsequent heating at 95 °C. A mixture of glucose (G5767, Sigma), laminaribiose (O-LAM2, Megazyme), laminaritriose (O-LAM3, Megazyme) and laminarihexaose (O-LAM6, Megazyme) was used as the thin-layer chromatography standards.

During enzymatic digestion of synthesized polymer, the reducing sugars released were measured using the DNS reagents (Micro Reducing Sugar Assay Kit, Abbkine). In brief, 175 µl of the diluted hydrolysis sample was mixed with 125 µl of DNS reagent. Standard from the kit at the concentration range of 0.2–0.6 mg ml^{-1} was used for the generation of the standard curve. Reaction mixtures were boiled in a water bath for 5 min. After cooling to room temperature in a water-ice bath, the absorbance of a 0.2 ml sample was measured at 540 nm.

Glycosyl linkage (methylation) analysis

The in vitro synthesized polymer by FKS1 was washed twice with buffer: 50 mM Tris-HCl pH 7.4, 33% glycerol, 1 mM EDTA, 0.2% CHAPS, 0.04% CHS, 20 mM KF and two times with deionized water. The washed sample was then dialysed into deionized water and freeze-dried. Methylation analysis was performed following a previous study[66]. The dried sample (approximately 1 mg) was dissolved in DMSO (500 µl). The methylation was performed in DMSO/NaOH with iodomethane for 1 h. The methylated products were hydrolysed with 2 M trifluoroacetic acid at 121 °C for 90 min, reduced by NaBD4 and acetylated with acetic anhydride at 100 °C for 2.5 h. The resulting partially methylated alditol acetates were analysed using the GC-MS system (6890A-5975C, Agilent Technology) equipped with an Agilent BPX70 chromatographic column (30 m × 0.25 mm × 0.25 µm; SGE). The temperature program was set as follows: 140 °C for 2 min, 140–230 °C at 3 °C per minute and 230 °C for 3 min.

Quantitative determination of β-1,3-glucan in cell walls

The levels of β-1,3-glucan were determined using the aniline blue assay, as previously described[23,26,67,68,69,70,71,72]. Testing strains were grown in YPD

media to exponential phase until OD_{600} was 0.5. The same amount of cells for each strain was collected by centrifugation at 5,000*g*. The collected cells were washed twice with TE buffer (10 mM Tris-HCl pH 8.0, and 1 mM EDTA). The pelleted cells were suspended with 0.5 ml TE buffer and mixed with 0.1 ml of 6 M NaOH. The mixture was incubated in an 80 °C water bath for 30 min to solubilize the glucan. Then, 2.1 ml of aniline blue (0.03% aniline blue, 0.18 M HCl and 0.49 M glycine-NaOH pH 9.5) was added to each sample. The samples were briefly vortexed and incubated at 50 °C for 30 min and an additional 30 min at 24 °C. Fluorescence was quantified using a fluorescence plate reader (CLARIOstar Plus, BMG Labtech) under an excitation wavelength of 400 nm and an emission wavelength of 460 nm with a cut-off of 455 nm. All samples were measured in triplicates.

Protein identification with mass spectrometry analysis

The protein samples were separated by SDS–PAGE gel and were stained with Coomassie. The gel band was manually excised and destained. The proteins were reduced with DTT, alkylated with IAA and digested with proteomics-grade trypsin in 20 mM ammonium bicarbonate[73]. The digestion was performed overnight at 37 °C and stopped by adding 2% formic acid. The peptides in gel were extracted using solution containing 2% formic acid and 67% acetonitrile. The peptides were vacuum-dried, resuspended in 0.1% formic acid, loaded onto the trap column nanoViper C18 (3 μm, 100 Å) and separated on an analytic column (Acclaim PepMap RSLC, 75 μm × 25 cm; C18 2 μm, 100 Å) using the EASY nLC 1200 HPLC system (Thermo Fisher). The elution gradient was 5–38% buffer A (0.1% formic acid and 80% acetonitrile) over 30 min. The mass spectrometry analysis was performed on a Q Exactive mass spectrometer (Thermo Fisher). The resulting data were converted to mgf file with ProteoWizard and analysed using the Mascot search engine for protein identification against a UniProt *Saccharomyces cerevisiae* database with a false discovery rate of less than 1%.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

EM density maps have been deposited in the Electron Microscopy Data Bank under the accession codes EMD-33154 (FKS1) and EMD-34115 (FKS1(S643P)). The coordinates have been deposited in the PDB under the accession codes 7XE4 (FKS1) and 7YUY (FKS1(S643P)). This study analysed several protein structures publicly available from the PDB under the accession codes 4HG6, 6WLB, 7SP7, 6YV8, 6iwr, 6h4m and 1qgq.

References

1. Wagener, J., Striegler, K. & Wagener, N. α- and β-1,3-Glucan synthesis and remodeling. *Curr. Top. Microbiol. Immunol.* **425**, 53–82 (2020).

2. Gow, N. A. R., Latge, J. P. & Munro, C. A. The fungal cell wall: structure, biosynthesis, and function. *Microbiol. Spect.* https://doi.org/10.1128/microbiolspec.FUNK-0035-2016 (2017).

3. Perfect, J. R. The antifungal pipeline: a reality check. *Nat. Rev. Drug Discov.* **16**, 603–616 (2017).

4. Hoenigl, M. et al. The antifungal pipeline: fosmanogepix, ibrexafungerp, olorofim, opelconazole, and rezafungin. *Drugs* **81**, 1703–1729 (2021).

5. Brown, G. D. et al. Hidden killers: human fungal infections. *Sci. Transl. Med.* **4**, 165rv113 (2012).

6. Hoenigl, M. Invasive fungal disease complicating COVID-19: when it rains it pours. *Clin. Infect. Dis.* **73**, e1645–e1648 (2021).

7. Perlin, D. S., Rautemaa-Richardson, R. & Alastruey-Izquierdo, A. The global problem of antifungal resistance: prevalence, mechanisms, and management. *Lancet Infect. Dis.* **17**, e383–e392 (2017).

8. Fisher, M. C., Hawkins, N. J., Sanglard, D. & Gurr, S. J. Worldwide emergence of resistance to antifungal drugs challenges human health

and food security. *Science* **360**, 739–742 (2018).

9. Perlin, D. S. Cell wall-modifying antifungal drugs. *Curr. Top. Microbiol. Immunol.* **425**, 255–275 (2020).

10. Inoue, S. B. et al. Characterization and gene cloning of 1,3-β-d-glucan synthase from *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **231**, 845–854 (1995).

11. Douglas, C. M. et al. The *Saccharomyces cerevisiae* FKS1 (ETG1) gene encodes an integral membrane protein which is a subunit of 1,3-β-d-glucan synthase. *Proc. Natl Acad. Sci. USA* **91**, 12907–12911 (1994).

12. Qadota, H. et al. Identification of yeast Rho1p GTPase as a regulatory subunit of 1,3-β-glucan synthase. *Science* **272**, 279–281 (1996).

13. Mazur, P. & Baginsky, W. In vitro activity of 1,3-β-d-glucan synthase requires the GTP-binding protein Rho1. *J. Biol. Chem.* **271**, 14604–14609 (1996).

14. Katiyar, S. K. et al. Fks1 and Fks2 are functionally redundant but differentially regulated in *Candida glabrata*: implications for echinocandin resistance. *Antimicrob. Agents Chemother.* **56**, 6304–6309 (2012).

15. Mazur, P. et al. Differential expression and function of two homologous subunits of yeast 1,3-β-d-glucan synthase. *Mol. Cell. Biol.* **15**, 5671–5681 (1995).

16. Thompson, J. R. et al. A glucan synthase FKS1 homolog in *Cryptococcus neoformans* is single copy and encodes an essential function. *J. Bacteriol.* **181**, 444–453 (1999).

17. Douglas, C. M. et al. Identification of the FKS1 gene of *Candida albicans* as the essential target of 1,3-β-d-glucan synthase inhibitors. *Antimicrob. Agents Chemother.* **41**, 2471–2479 (1997).

18. Dichtl, K. et al. *Aspergillus fumigatus* devoid of cell wall β-1,3-glucan is viable, massively sheds galactomannan and is killed by septum

formation inhibitors. *Mol. Microbiol.* **95**, 458–471 (2015).

19. Chen, S. C., Slavin, M. A. & Sorrell, T. C. Echinocandin antifungal drugs in fungal infections: a comparison. *Drugs* **71**, 11–41 (2011).

20. Perlin, D. S. Mechanisms of echinocandin antifungal drug resistance. *Ann. N.Y. Acad. Sci.* **1354**, 1–11 (2015).

21. Alexander, B. D. et al. Increasing echinocandin resistance in *Candida glabrata*: clinical failure correlates with presence of FKS mutations and elevated minimum inhibitory concentrations. *Clin. Infect. Dis.* **56**, 1724–1732 (2013).

22. Johnson, M. E., Katiyar, S. K. & Edlind, T. D. New Fks hot spot for acquired echinocandin resistance in *Saccharomyces cerevisiae* and its contribution to intrinsic resistance of *Scedosporium* species. *Antimicrob. Agents Chemother.* **55**, 3774–3781 (2011).

23. Park, S. et al. Specific substitutions in the echinocandin target Fks1p account for reduced susceptibility of rare laboratory and clinical *Candida* sp. isolates. *Antimicrob. Agents Chemother.* **49**, 3264–3273 (2005).

24. Johnson, M. E. & Edlind, T. D. Topological and mutational analysis of *Saccharomyces cerevisiae* Fks1. *Eukaryot. Cell* **11**, 952–960 (2012).

25. Chhetri, A. et al. Length specificity and polymerization mechanism of (1,3)-β-d-glucan synthase in fungal cell wall biosynthesis. *Biochemistry* **59**, 682–693 (2020).

26. Suwunnakorn, S., Wakabayashi, H., Kordalewska, M., Perlin, D. S. & Rustchenko, E. FKS2 and FKS3 genes of opportunistic human pathogen *Candida albicans* influence echinocandin susceptibility. *Antimicrob. Agents Chemother.* **62**, e02299-17 (2018).

27. Beauvais, A. et al. Glucan synthase complex of *Aspergillus fumigatus*. *J. Bacteriol.* **183**, 2273–2279 (2001).

28. Perlin, D. S. Echinocandin resistance, susceptibility testing and prophylaxis: implications for patient management. *Drugs* **74**, 1573–1585 (2014).

29. Mullins, J. T. Regulatory mechanisms of β-glucan synthases in bacteria, fungi, and plants. *Physiol. Plant.* **78**, 309–314 (1990).

30. Frost, D. J., Brandt, K., Capobianco, J. & Goldman, R. Characterization of (1,3)-β-glucan synthase in *Candida albicans*: microsomal assay from the yeast or mycelial morphological forms and a permeabilized whole-cell assay. *Microbiology* **140**, 2239–2246 (1994).

31. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).

32. Morgan, J. L., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat. Struct. Mol. Biol.* **21**, 489–496 (2014).

33. Purushotham, P., Ho, R. & Zimmer, J. Architecture of a catalytically active homotrimeric plant cellulose synthase complex. *Science* **369**, 1089–1094 (2020).

34. Maloney, F. P. et al. Structure, substrate recognition and initiation of hyaluronan synthase. *Nature* **604**, 195–201 (2022).

35. Gandini, R. et al. A transmembrane crenarchaeal mannosyltransferase is involved in N-glycan biosynthesis and displays an unexpected minimal cellulose-synthase-like fold. *J. Mol. Biol.* **432**, 4658–4672 (2020).

36. Morgan, J. L., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181–186 (2013).

37. Stasinopoulos, S. J., Fisher, P. R., Stone, B. A. & Stanisich, V. A. Detection of two loci involved in (1→3)-β-glucan (curdlan) biosynthesis by *Agrobacterium* sp. ATCC31749, and comparative

sequence analysis of the putative curdlan synthase gene. *Glycobiology* **9**, 31–41 (1999).

38. Bi, Y., Hubbard, C., Purushotham, P. & Zimmer, J. Insights into the structure and function of membrane-integrated processive glycosyltransferases. *Curr. Opin. Struct. Biol.* **34**, 78–86 (2015).

39. Wu, X. D. & Rapoport, T. A. Translocation of proteins through a distorted lipid bilayer. *Trends Cell Biol.* **31**, 473–484 (2021).

40. Carolus, H. et al. Genome-wide analysis of experimentally evolved *Candida auris* reveals multiple novel mechanisms of multidrug resistance. *mBio* **12**, e03333-20 (2021).

41. Jimenez-Ortigosa, C., Perez, W. B., Angulo, D., Borroto-Esoda, K. & Perlin, D. S. De novo acquisition of resistance to SCY-078 in *Candida glabrata* involves FKS mutations that both overlap and are distinct from those conferring echinocandin resistance. *Antimicrob. Agents Chemother.* **61**, e00833-17 (2017).

42. Muller, A. et al. Daptomycin inhibits cell envelope synthesis by interfering with fluid membrane microdomains. *Proc. Natl Acad. Sci. USA* **113**, E7077–E7086 (2016).

43. Lewis, K. The science of antibiotic discovery. *Cell* **181**, 29–45 (2020).

44. Bessueille, L. et al. Plasma membrane microdomains from hybrid aspen cells are involved in cell wall polysaccharide biosynthesis. *Biochem. J.* **420**, 93–103 (2009).

45. Briolay, A. et al. Cell wall polysaccharide synthases are located in detergent-resistant membrane microdomains in oomycetes. *Appl. Environ. Microbiol.* **75**, 1938–1949 (2009).

46. Healey, K. R., Katiyar, S. K., Raj, S. & Edlind, T. D. CRS-MIS in *Candida glabrata*: sphingolipids modulate echinocandin–Fks interaction. *Mol. Microbiol.* **86**, 303–313 (2012).

47. Satish, S. et al. Stress-induced changes in the lipid microenvironment of β-(1,3)-d-glucan synthase cause clinically important echinocandin resistance in *Aspergillus fumigatus*. *mBio* **10**, e00779-19 (2019).

48. Funakoshi, M. & Hochstrasser, M. Small epitope-linker modules for PCR-based C-terminal tagging in *Saccharomyces cerevisiae*. *Yeast* **26**, 185–192 (2009).

49. Toulmay, A. & Schneiter, R. A two-step method for the introduction of single or multiple defined point mutations into the genome of *Saccharomyces cerevisiae*. *Yeast* **23**, 825–831 (2006).

50. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

51. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

52. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).

53. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).

55. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

57. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

58. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).

59. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).

60. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

61. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).

62. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

63. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

64. Zhang, Y. H., Cui, J., Lynd, L. R. & Kuang, L. R. A transition from cellulose swelling to cellulose dissolution by *o*-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7**, 644–648 (2006).

65. McIntosh, M., Stone, B. A. & Stanisich, V. A. Curdlan and other bacterial (1→3)-β-d-glucans. *Appl. Microbiol. Biot.* **68**, 163–173 (2005).

66. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217 (1984).

67. Shedletzky, E., Unger, C. & Delmer, D. P. A microtiter-based fluorescence assay for (1,3)-β-glucan synthases. *Anal. Biochem.* **249**, 88–93 (1997).

68. Kahn, J. N., Hsu, M. J., Racine, F., Giacobbe, R. & Motyl, M. Caspofungin susceptibility in *Aspergillus* and non-*Aspergillus* molds: inhibition of glucan synthase and reduction of β-d-1,3 glucan levels in culture. *Antimicrob. Agents Chemother.* **50**, 2214–2216 (2006).

69. Sekiya-Kawasaki, M. et al. Dissection of upstream regulatory components of the Rho1p effector, 1,3-β-glucan synthase, in *Saccharomyces cerevisiae*. *Genetics* **162**, 663–676 (2002).

70. Watanabe, D., Abe, M. & Ohya, Y. Yeast Lrg1p acts as a specialized RhoGAP regulating 1,3-β-glucan synthesis. *Yeast* **18**, 943–951 (2001).

71. Lee, K. K. et al. Yeast species-specific, differential inhibition of β-1,3-glucan synthesis by poacic acid and caspofungin. *Cell Surf.* **3**, 12–25 (2018).

72. Anjos, J. et al. β(1,3)-Glucan synthase complex from *Alternaria infectoria*, a rare dematiaceous human pathogen. *Med. Mycol.* **50**, 716–725 (2012).

73. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).

Acknowledgements

We thank all staff members at the Cryo-EM Center, Southern University of Science and Technology for their assistance in data collection; and J. Zhao for advice on thin-layer chromatography. This work was supported by the National Natural Science Foundation of China (92053112 and 31971148 to H.Y., 32100575 to Min Zhang and 82188101 to Mingjie Zhang), ShenZhen Talent Program KQTD20210811090115021 (to Mingjie Zhang and X.L.), the Fundamental Research Funds for the Central Universities (5003510056

to H.Y. and 5003510112 to Min Zhang) and Program of HUST Academic Frontier Youth Team (2018QYTD02 to H.Y.).

Author information

Author notes

1. These authors contributed equally: Xinlin Hu, Ping Yang

Authors and Affiliations

1. Department of Biochemistry and Molecular Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

   Xinlin Hu, Ping Yang, Changdong Chai, Jia Liu, Huanhuan Sun, Yanan Wu, Min Zhang & Hongjun Yu

2. Department of Pathogen Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

   Xinlin Hu, Jia Liu & Min Zhang

3. School of Life Sciences, Southern University of Science and Technology, Shenzhen, China

   Mingjie Zhang & Xiaotian Liu

4. Greater Bay Biomedical Innocenter, Shenzhen Bay Laboratory, Shenzhen, China

   Mingjie Zhang

5. Cell Architecture Research Center, Huazhong University of Science and Technology, Wuhan, China

   Hongjun Yu

Contributions

X.H., P.Y., Min Zhang, X.L. and H.Y. prepared the sample, collected the data and solved the structures. X.H., P.Y. and Min Zhang performed the functional experiments, assisted by C.C., J.L., H.S. and Y.W. Mingjie Zhang, Min Zhang, X.L. and H.Y. designed the experiments, analysed the data and wrote the manuscript.

Corresponding authors

Correspondence to Min Zhang, Xiaotian Liu or Hongjun Yu.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Vincent Bulone and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Sample preparation, functional characterization and in vitro product synthesis.

a, Elution profile of Gel-filtration chromatography (Superose 6 Increase 10/300 GL) of FKS1 purified in detergent GDN. Green dashed box marks the fractions pooled for cryoEM analysis. **b**, SDS-PAGE (left) and western blot (right) analysis of the fractions of the monodisperse elution peak in (**a**). Representative gel and western-blot image among 3 replicates are shown. **c**, Elution profile of Gel-filtration chromatography (Superose 6 Increase 10/300 GL) of FKS1 S643 purified in detergent GDN. Green dashed box marks the fractions pooled for cryoEM analysis and product synthesis. **d**, SDS-PAGE analysis of the fractions of the monodisperse elution peak in (**c**). A representative gel among 3 replicates is shown. **e**, Elution profile of Gel-filtration chromatography (Superdex 200 Increase 10/300 GL) of Rho1. Green dashed box marks the fractions pooled for activity assay and product synthesis. **f**, SDS-PAGE analysis of the fractions of the monodisperse elution peak in (**e**). Two peaks in (**e**) shows similar three-band pattern (**f**), and all three bands were identified as Rho1 by Mass Spec analysis. A representative gel among 3 replicates is shown. **g**, Western blot analyses of wild-type FKS1 and different FKS1 variants purified in detergent CHAPS. A representative image among 3 replicates is shown. **h**, SDS-PAGE analysis of GDN-purified flag-tagged FKS1 variants, which were used for the assays of activity (**i-j**) and product synthesis (**k**). First lane, FKS1 WT; second lane, FKS1 S643P (the echinocandin-resistant mutant); third lane, purified FKS1 S643P is immunodepleted by anti-flag beads; fourth lane, FKS1 S643P/K1261A. A representative gel among 3 replicates is shown. For (**b**,**d**,**f**,**g**,**h**), their full scans are provided in Supplementary Fig. 1. **i**, In vitro activity of FKS1 variants purified in GDN. The activity was assayed by monitoring UDP generated (1-h reaction). Data are mean ± SEM, n = 3 independent experiments. **j**–**k**, In vitro product synthesis assay by FKS1 variants purified in GDN. The assay was performed either by staining the synthesized products with dye aniline blue (**j**; 12-h reaction) or by visualization of the products synthesized (**k**; 48-h reaction). These experiments were repeated three times with similar results. **l**, Donor specificity of product synthesis by GDN-purified FKS1 S643P. Water-insoluble polymers (indicated by white arrow) appear only in the presence of UDP-Glc instead of UDP-GlcNAc. This experiment was repeated three times with similar results. **m**, The effects of Mg^{2+} on the catalytic activity of GDN-purified FKS1 S643P. Activity was assayed in the presence of 1 mM EDTA or 200 μM Mg^{2+}. Data are mean ± SEM, n = 3 independent

experiments. Inset, product synthesis in the presence of 1 mM EDTA or 200 μM Mg^{2+}. This experiment was repeated three times with similar results.

Extended Data Fig. 2 Enzymatic hydrolysis and glycosyl linkage (methylation) analysis of FKS1-synthesized products.

a, Quantification of the released reducing sugars during enzymatic digestion of the water-insoluble polymer synthesized by GDN-purified FKS1 S643P. Two specific glucanases were used: specific endo-1,3-β-glucanase and endo-1,4-β-glucanase. Data are mean ± SEM, n = 3 independent experiments. **b**–**c**, The hydrolysis specificity of endo-1,3-β-Glucanase and endo-1,4-β-Glucanase was tested against standard polysaccharide β-1,3-glucan curdlan (**b**) and cellulose (**c**). A mixture of glucose (G1), laminaribiose (G2), laminaritriose (G3) and laminarihexaose (G6) was used as the standards (lane M). **d**, Mass spectrum of GC peak (11.64min) in (Fig. 1f) reveals the identity of this PMAA as 1,3,5-tri-O-acetyl-2,4,6-tri-O-methyl glucitol, confirming 1,3-Glc*p* linkage.

Extended Data Fig. 3 Cryo-EM structural determination of *S. cerevisiae* FKS1 and structural features of FKS1 cytoplasmic region.

a, A representative cryo-EM micrograph of FKS1 from 11606 collected micrographs. **b**, Representative 2D class averages. **c**, Flow chart of cryo-EM data acquisition and data processing of FKS1. See Methods for more details. **d**, The gold-standard Fourier shell correlation (FSC) curve of the reconstructed map. **e**, The FSC curve between model and map, which is calculated by PHENIX.mtriage[58]. **f**, Cutaway views of the angular distribution of all particles used in the final 3D reconstruction. **g**, Local resolution distribution of the final cryo-EM map of FKS1, calculated with ResMap[54]. **h**, Topological diagram of FKS1 GT-domain. Flexible segments invisible in the 3D reconstruction are depicted by dashed lines. **i**, Interactions between AC and GT domain. Labeled are central β strands and the structural elements involved in the extensive contacting interface: α2-5, α8-11, loops Lα9-α10 and Lα11-α12 of AC domain; α31, α34-35, loops Lβ5-β6, Lβ7-β8; Lβ8-β9 of GT domain). The FKS1 R319, a previously

identified essential residue for FKS1[24], is marked by its Cα atom as sphere and is embedded in this interface. **j**, Zoomed in view of the boxed region in (**i**), showing the interaction between R319 and N1087 to maintain the interface between AC and GT domain.

Extended Data Fig. 4 Fit of cryo-EM map with FKS1 model in example regions.

a, Cryo-EM densities of TM1-8, 11–17. **b**-**c**, Cryo-EM densities of the continuous central β-sheet of 9 strands featured in FKS1 GT domain. **d**, Cryo-EM density of the N-glycan on N1849. **e**, Cryo-EM densities of active site. **f**, Cutaway view of the cryo-EM densities near TM9 and TM10. The map is segmented and colored as in Fig. 2a. The densities corresponding to TM helices TM7-12, TM14-16 are denoted by numbers. Only TM9-10 main chain can be traced from this map, though they exhibit weaker densities than other TM helices. **g**–**h**, Two views of FKS1 model with bound lipids. FKS1 model is colored as Fig. 1h and ordered lipids are shown as yellow spheres. The cryo-EM densities of the lipids are illustrated in the dashed boxes. The position and the density of the phospholipid embedded within the transmembrane domain is shown in the solid box.

Extended Data Fig. 5 Structure and sequence comparison of the conserved cellulose-synthase-like folds between FKS1 and BcsA.

a, The conserved fold of FKS1 (residues 615–1510) was superposed with that of BcsA (PDB ID: 4P00; residues 63–584), with an RMSD value of 4.3 Å over 308 aligned residues. BcsA is colored in grey. FKS1 core is colored in orange while FKS1-exclusive elements in its GT domain are in color-coded display as labeled in (**a**) and (**b**). Corresponding TM helices are denoted as labels while TM9 and TM10 are exclusive for FKS1. **b**, Close-up view of the superimposed GT domains in (**a**). The central β-sheet featured in GT domain gains extension in FKS1. The extended stands β6 and β7 in FKS1 and the elongated loop Lβ8-β9 contribute to a substantial portion of the interacting interface with AC domain in FKS1 (Fig. 1j). Moreover, two FKS1-specific sub-regions inserts to two sides of β3, the first strand of central β-sheet. The sub-region preceding β3 (residues 719–840) contains six helices and a pair of antiparallel β-strands (β1, β2); the sub-region

ensuing β3 (residues 848–998) is α-helical, composed of 9 helices. These two sub-regions interact to pack the central β-sheet from one side. **c**, Back view of the superimposed transmembrane domains. **d**, Close-up view of superimposed membrane-cytosol interfaces. Interface helix IF1 is preserved in FKS1, which shows dramatic rearrangement relative to that of BcsA. Interface helix IF2 connecting a beta-sheet and transmembrane domain is preserved in FKS1 but disordered in the structure, which is also indicated in (**c**). Interface helix IF3 is gone in FKS1 and it is replaced by transmembrane helices TM9 and TM10 in FKS1. **e**, Structure-based sequence alignment of the conserved cellulose-synthase-like folds between FKS1 and BcsA. The alignment was first generated using PROMALS3D and then illustrated using ESPript3 server. Residues and motifs with important functions are labeled. Shown above and below the alignment are secondary structural elements derived FKS1 structure (blue symbol) and BcsA structure (PDB ID: 4P00) (orange symbol), respectively. The symbols are TM helices (filled bar), helices (empty bar) and β strands (arrow).

Extended Data Fig. 6 Overlay of the GT catalytic domains of the structural matches to FKS1 by DALI analysis.

The analyzed GT catalytic domains are overlaid and shown separately as cartoon. Residues involved in substrate binding or catalysis are indicated as sticks. The bound nucleotide is shown as thin magenta sticks. The conserved residues are highlighted as yellow when using the active site of bacterial cellulose synthase BcsA (**a**) as reference. **a**–**f**, Overlay of GT catalytic domains of membrane-bound glycosyltransferases: bacterial cellulose synthase BcsA (**a**; PDB ID: 4hg6), β-1,3-glucan synthase FKS1 (**b**), plant cellulose synthase CesA (**c**; PDB ID: 6wlb), virus hyaluronan synthase (**d**; PDB ID: 7sp7), Agrobacterium curdlan synthase (**e**; model from AlphaFold Database; Uniprot ID: Q9X2V0), archaeal mannosyltransferase PcManGT (**f**; PDB ID: 6YV8). The panels outlined by blue box (**a**–**e**) are membrane-bound polysaccharide synthases containing cellulose-synthase-like fold as shown in Fig. 2f–m. PcManGT (panel **f**) contains half membrane tunnel of BcsA, and it was proposed to have a minimal cellulose-synthase-like fold[35]. **g**–**i**, Overlay of GT catalytic domains of soluble glycosyltransferases: human GalNAc-T7 (**g**; PDB ID: 6iwr), *S. aureus* TarP (**h**; PDB ID:6h4m), *B. subtilis* SpsA (**i**; PDB ID: 1qgq).

Extended Data Fig. 7 Cryo-EM analysis of FKS1 S643P, a drug-resistant mutant of the enzyme.

a, A representative cryo-EM micrograph of FKS1 S643P from 9623 collected micrographs. **b**, Representative 2D class averages. **c**, Flow chart of cryo-EM data acquisition and data processing of FKS1 S643P. See Methods for more details. **d**, The gold-standard Fourier shell correlation (FSC) curve of the reconstructed map. **e**, The FSC curve between model and map, which is calculated by PHENIX.mtriage[58]. **f**, Cutaway views of the angular distribution of all particles used in the final 3D reconstruction. **g**, Local resolution distribution of the final cryo-EM map of FKS1 S643P, calculated with ResMap. **h**, Fit of cryo-EM map with FKS1 S643P model in example regions.

Extended Data Fig. 8 Analysis of caspofungin-resistant phenotype and chemical structure of echinocandin drugs.

a, Analysis of caspofungin-resistant phenotype of *S. cerevisiae* strain with indicated FKS1 mutations. The concentrations of caspofungin tested are indicated on the top. **b**, Chemical structure of four types of echinocandin drugs. They all feature a lipid tail as highlighted by the box. **c**, Chemical structure of the lipid alkyl chains (left panel) and phospholipid (right panel) identified from FKS1 structure.

Extended Data Fig. 9 Sequence conservation of fungi FKS illustrated on *S. cerevisiae* FKS1 structure.

a—**c**, FKS1 structure is colored according to level of sequence conservation (red, high; blue, low). **a**–**b**, Back and front view parallel to the membrane, in cartoon display. **c**, The same view as (**b**) with surface representation.

Extended Data Table 1 Cryo-EM data collection, refinement and validation statistics

Supplementary information

Supplementary Figures

This file contains the uncropped blots (Supplementary Fig. 1) and the multiple sequence alignment of FKS orthologs from various pathogenic fungi and *Saccharomyces cerevisiae* (Supplementary Fig. 2).

Reporting Summary

Rights and permissions

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Reprints and Permissions

About this article

Cite this article

Hu, X., Yang, P., Chai, C. *et al.* Structural and mechanistic insights into fungal β-1,3-glucan synthase FKS1. *Nature* **616**, 190–198 (2023). https://doi.org/10.1038/s41586-023-05856-5

- Received: 29 March 2022
- Accepted: 16 February 2023
- Published: 22 March 2023
- Issue Date: 06 April 2023
- DOI: https://doi.org/10.1038/s41586-023-05856-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05856-5